

TARTU ÜLIKOOL
SOTSIAALTEADUSTE VALDKOND
NARVA KOLLEDŽ
ÕPPEKAVA „INFOTEHNOLOOGILISTE SÜSTEEMIDE ARENDUS“

Olga Gandshu
LASTE KÕNE KORPUSE „RusLAPSED“ MÄRGISTUSE JA OTSINGU
INFOSÜSTEEM
Lõputöö
Juhendaja assistent Andre Säask

NARVA 2020

Olen koostanud töö iseseisvalt. Kõik töö koostamisel kasutatud teiste autorite tööd, põhimõttelised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

.....

Töö autori allkiri ja kuupäev

SISUKORD

SISUKORD.....	2
SISSEJUHATUS	4
I. INFOSÜSTEEMI VAJADUS	5
1. Laste kõne tekste uurimise probleem	5
2. Keelekorpusega tutvumine	6
3. RusLAPSED infosüsteemi ootused	8
3.1 Infosüsteemi kasutajad	9
3.2 Infosüsteemi osad	10
4. Olemasolevate lahenduste uuring	14
4.1 Tallinna Ülikooli eesti vahekeele korpus (EVKK)	14
4.2 Corpus of Russian Student Texts (CoRST)	18
4.3 Vene keele rahvuskorpus (Национальный корпус русского языка)	21
4.4 Korpuste võrdlemine	26
5. Eeluurimuse kokkuvõte	28
II. ARENDATAVA SÜSTEEMI ANALÜÜS	29
1. Projekti haldamise vahendid	29
2. Nõuded	29
2.1 Funktsionaalsed nõuded	30
2.2 Mittefunktsionaalsed nõuded	31
3. Kasutusloode koostamine ja prototüüpimine	32
3.1 Veebilehe päis ja menüü	32
3.2 Avaleht	33
3.3 Sisselogimine	34
3.4 Isiklik konto	35
3.5 Teksti lisamine	37
3.6 Otsing	44
4. Tehnoloogiate valik	47
4.1 Angular 8	47
4.2 ASP.NET Core Web API	48
4.3 MS SQL Express ja Entity Framework Core	48
4.4 Bootstrap	49
III. INFOSÜSTEEMI PROTOTÜÜBI LOOMINE	50
1. Rakenduse arhitektuurne lahendus	50
1.1. Projekti üldehitus	50
1.2. Infosüsteemi kliendiosa	52

2.	Andmemudel	54
2.1.	Olemite metaandmed	54
2.2.	Uudiste ja korpuse info tabelid	55
2.3.	Kasutajate ja kasutaja sõnumite tabelid	56
2.4.	Klassifikaatorite ja žanrite tabelid.....	58
2.5.	Peamine andmemudeli osa teksti tabeliga	59
3.	Kliendi ja serveri osade suhtlemine	66
4.	RusLAPSED veebilehe disain	68
4.1.	Veebilehe disaini muutused	68
5.	Viited lõputöö olulistele lisadele.....	70
IV.	JÄRGMISED SAMMUD INFOSÜSTEEMI ARENDAMISEKS	71
1.	Töö sõnavormiga.....	71
2.	Töö teksti kontrollimisega.....	71
3.	Teksti otsing sõna järgi	71
4.	Kasutajaliidese mitmekeelsus	72
5.	Muu funktsionaalsus	72
	KOKKUVÕTE	73
	SUMMARY	74
	KIRJANDUS	76
	LISAD	79
	LISA 1. CoRST ja VKRK korpuste kasutajaliideste näited	80
	LISA 2. Projekti haldamise vahendite vaated	83
	LISA 3. Teksti metaandmete tabelid.....	85
	LISA 6. RusLAPSED andmemuudel	91
	LISA 5. RusLAPSED kasutajaliidese pildid.....	92
	LISA 6. Korpuse RusLAPSED teksti otsingu filtritega eskiis.....	104
	Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks...	105

SISSEJUHATUS

Lõputöö teemaks on ühe erialase infosüsteemi, laste kõne korpuse RusLAPSED, esimese arenduse etapi ellu viimine. See töö on suure infosüsteemi arenduse algus, mis on mõeldud laste kõne tekstide lisamiseks andmebaasi ning nende tekstide sees otsimiseks sõnade ja erinevate tunnuste järgi.

Lõputöö teema oli pakutud lõputöö juhendaja poolt, kelle poole omakorda pöördus TÜ Narva Kolledži vene keele lektoraadi esindaja. Koos oma tudengitega ta tegeleb laste kõne uurimisega. Neid huvitab ainult kõne, mille autoriks on lapsed, kes elavad Eestis ja kelle emakeeleks on vene keel. Uurimiste jaoks soovitakse kasutada teksti märgistuse ja otsingu infosüsteemi ehk keele korpust.

Lõputöö probleem seisneb selles, et Eestis puuduvad venekeelsed korpused, kus on olemas laste kõne spetsiifiliste tunnuste märgistuse ja otsingu võimalused. See teema oli autori jaoks väga huvitav, sest sisaldas reaalse tarkvara arendust, mis võib olla suureks abiks vene keele lektoraadi tudengitele laste kõne ja selle arengu uurimises.

Antud töö sisaldab uuringu infosüsteemi vajadusest ning ülevaate sarnastest vahenditest teksti kogumiseks ja analüüsimiseks; lisaks sellele nõuete analüüsi, töövahendite valiku põhjendust ning kõikide arenduse etappide kirjeldusi.

Lõputöö põhiülesanne on luua infosüsteemi algversiooni, mida oleks võimalik edasi arendada teiste arendajate poolt järgmistel aastatel. „Arendaja“ sõna all mõeldakse siin mitte ainult tarkvara arendajaid, vaid ka vene keele lektoraadi töötajaid ja üliõpilasi, kes edaspidi hakkavad tegelema tekstide lisamisega, ning sõnade märgistusega.

I. INFOSÜSTEEMI VAJADUS

1. Laste kõne tekste uurimise probleem

TÜ Narva Kolledži vene keele lektoraat koos oma tudengitega tegeleb laste kõne uurimise ja analüüsimisega. Kõne uurimiseks tudengid kasutavad nende või nende eelkäijate poolt kogutud tekstid. Kõikide tekstide, dialoogide ja polüloogide autoriks on lapsed. Kogutud materjalid võivad sisaldada ka täiskasvanu inimeste kõnesid, kuid see ei osale sõnade märgistuse analüüsis. Sisu tekstide koostamiseks tudengid koguvad erinevatest allikatest: veebipõhisest kirjavahetusest, isiklikest vestlustest lastega või dialoogide/polüloogide üleskirjutamisest laste mängu ajal. Ükskõik millise lapse kõne üleskirjutamine toimub ainult siis kui lapsevanem või eestkostja on andnud selleks luba.

Erinevate tunnuste järgi laste kõne uurimiseks ei piisa ainult teksti enda olemasolust, seepärast lisatakse tekstile erinevaid metaandmeid. Tekstile lisatakse info teksti koguja kohta (inimene, kes kirjutas üles laste sõnad); lapse või laste kohta, kui neid oli mitu; ning tingimuste kohta üles kirjutamise ajal. Informatsioon lapse kohta sisaldab üldiseid andmeid pere, emakeele, õppimise tingimuste ja lapse vanuse kohta ning selliseid andmeid, mis on nõus jagada kogujaga lapsevanem või eestkostja. Reaalsete laste nimesid ei salvestata, iga lapse jaoks koguja mõtleb välja varjunime. Metaandmed omavad tähtsust edaspidise uuringu jaoks ja erinevate laste gruppide võrdluse jaoks.

Aastate jooksul on kogutud suur kogus tekste (umbes 840), mida on üsna raske kasutada, kuna vajalike uurimismaterjalide leidmine võib võtta väga palju aega. Näiteks 5-6-aastaste laste kõne uurimiseks on vaja käsitsi sorteerida saadaolevaid materjale ja valida ainult neid, mis sobivad uuringuks. Samuti on keeruline saadaolevate andmete struktureerimine, kuna tekstide puhul pole oluline ainult osalenud lapse vanus, vaid ka palju muid tegureid.

Selliste materjalidega töötamiseks on palju tööriistu, näiteks, erinevate keelte keelekorpused, millel võivad olla mitmesugused eesmärgid tekstide uurimiseks ja analüüsimiseks. Kui piirata olemasolevate keelekorpusete arv TÜ Narva kolledži tudengite vajadustega, nimelt laste venekeelse kõnekorpusega, siis pole nende arv suur ja probleem on selles, et need tööriistad sisaldavad peamiselt Vene Föderatsioonis kogutud tekste ja neid ei saa rakendada Eestis elavate laste kõne uurimisel. Eestis elavatel venekeelsetel

lastel on unikaalne kultuuriline ja keeleline keskkond, ning nende kõne areng erineb teistest, mis omakorda pakub huvi uurimistööde jaoks.

Idee luua venekeelsete laste kõnekorpus tekkis sellise tööriista puudumise tõttu Eestis. Kõnekorpus RusLAPSED võib olla heaks aluseks Eestis elavate venekeelsete laste kõne edaspidiseks uurimiseks. Infosüsteemi saavad kasutada mitte ainult TÕ Narva Kolledžis õppivad tudengid, vaid ka kõik inimesed, keda see huvitab.

2. Keelekorpusega tutvumine

Keelekorpuse mõiste erineb natuke erinevates allikates. Üks allikas määratleb seda lihtsalt suulise ja kirjaliku kõne kogumina, teine võib määratleda seda elektrooniliste tekstide või kõnede kogumina. Keelekorpuse mõiste sõltub selle määratlemise ajast. (Arvuti maailm veebileht; Muischnek, Orav jt 2003: 9-10)

„Keeleteaduses on sõna korpus all enne arvutite kasutuselevõttu tavaliselt mõeldud keeleainese kogumikku, mida kasutatakse uurimistöös materjalina (esineb see siis kartoteegi, lindikogu vms. kujul) vastandina autori enda intuitsioonil põhinevatele üldistusele“ (Muischnek, Orav jt 2003: 9).

Tehnoloogia arenguga keelekorpused on läinud üle elektroonilisele vormile, mis võimaldab olemasolevaid materjale mugavamalt kasutada. Kadri Muischnek ja Heili Orav andsid oma artiklis „Eesti keele korpused ja arvutileksikonid — mis on olemas ja mida veel vaja on“ keelekorpusele määratluse:

„Korpus on elektrooniline keele (teksti või kõne) kogum, mille alusel saab:

- analüüsida keelt, et tema omadusi kindlaks teha;
- treenida mingit arvutiprogrammi, et kohandada teda tööks tekstidega teatud piiritletud olukorras;
- empiirilisel kontrollida keele kohta käivat teooriat;
- testida keeletehnoloogilist võtet või rakendust, et selgitada, kuidas ta töötab praktikas“ (Arvuti maailm veebileht).

Esmaste elektrooniliste teksti keelekorpustena loetakse Browni ja Lancaster-Oslo/Bergeni tekstikorpused (Muischnek, Orav jt 2003: 9-10).

Browni tekstikorpus – on tavalise ameerika inglise keele korpus, mis loodi Portugali Brown University W.N. Francis ja H. Kucera (koostajad) 1964. aastal. Korpuse maht – natuke üle miljoni sõnu kirjanduse tekstidest, kirjutatud 1961. aastal ja varem ameerika inglise keeles. Korpus jaguneb mitmeks osaks – teemade järgi. Iga tekst sisaldab 2000 sõna ja kui tegu on suurema mahuga, siis on see osaliselt sisestatud. Selle korpuse põhjal loodi veel kaks korpust: LOB corpus (British English) ja Kolhapur Corpus (Indian English). (Brown'i korpuse ja Brown'i korpuse kasutusjuhendi veebilehed)

Lancaster-Oslo/Bergeni või LOB tekstikorpus – on 1961. aastal briti inglise keeles kirjutatud kirjandustekstide põhjal miljoni sõna korpus. LOB tekstikorpus töötati välja järgmiste organisatsioonide koostöö tulemusel: University of Lancaster, the University of Oslo, ja the Norwegian Computing Centre for the Humanities at Bergen. Korpus koostati 8 aasta jooksul, aastatel 1970–1978. See korpus on Brown'i teksti korpuse analoog ning sisaldab igas tekstis ka umbes 2000 sõna. Seda on tehtud selleks, et lihtsustada kahe korpuse üheaegne kasutamine. (LOB'i korpuse kasutusjuhend)

Praegusel ajal on olemas suur hulk keelte korpuseid, mis erinevad mitte ainult sisu keele järgi, vaid ka uurimise eesmärkide ja märgistusdetailsuse järgi. Kaasaegsed korpused võivad sisaldada kirjanduslikke tekste, kõne-, heli- ja videosalvestusi, kirjutatud või räägitud tekste, spontaanseid kõnetekste, dialooge – kõike materjale, mis võivad uurijatele huvi pakkuda.

Erinevates korpustes tehakse sõnade märgistamine erineval viisil, mõnel juhul on ripploendid võimalusega valida vajalikke andmeid ja mõnel juhul on oma märgistuskeel. Oma märgistuskeel on üsna tavaline viis tekstisõnade märgistamiseks erinevates korpustes. Samuti paljudel juhtudel tuleb seda keelt päringus kasutada, sisestades selle märgistuskeele või metakeele elemendid koos sõnaga ühes päringu reas. Autori arvates muudab see materjaliga töötamise oluliselt keerukamaks, kuna see nõuab selle märgistuskeele selgeksõppimist, selleks et saaks märgistuskeelt kasutada.

Keelekorpused erinevad ka mahult, mis tähendab sisalduvate tekstide, sõnavormide või helisalvestiste arvu. Esimeste korpuste eelnevalt toodud näidetel on tänapäevaste süsteemidega võrreldes üsna väikese mahuga.

Raamatus „Eesti keele tehnoloogilised ressursid ja vahendid” maailma kõige mahukamaks keelekorpuseks nimetakse Mannheimi Saksa Keele Instituudis koostatud

Korpora geschriebener Gegenwartssprache, mis sisaldab 2 miljardit sõnakasutust (Muischnek, Orav jt 2003: 11). See oli seis 2003. aastal, 2020. aasta seisuga sisaldab see korpus 49,6 miljardit sõnakasutust ja seda peetakse maailma suurimaks saksakeelsete kirjalike korpuste koguks (Saksa keele instituudi korpus). Autor ei tea, kas tänapäeval on mõni mahukam keelekorpus olemas.

Eestis kõige mahukamaks keelekorpuseks loetakse internetikorpust etTenTen (Eesti keele koondkorpuse veebileht). Keelekorpus sisaldab 270 miljonit sõnakasutust 686 000 eestikeelsest veebilehest (etTenTen'i veebileht).

Veel üks suur Eesti ressurss – Eesti keele koondkorpus, mis sisaldab paljusid teisi korpuse kirjaliku keele täistekstidega. Korpuse ligikaudne maht on 240 miljonit sõnakasutust. (Eesti keele koondkorpuse veebileht)

Korpuse maht ei ole muidugi selle peamine näitaja. Igal korpusel on oma eesmärk ning see on välja töötatud kasutajate erinevateks vajadusteks. Keelekorpus võib olla väike, kuid sellegipoolest võib temast olla teadusuuringutes palju kasu.

3. RusLAPSED infosüsteemi ootused

RusLAPSED korpuse esitatavate nõuete väljaselgitamiseks korraldati kohtumised TÜ Narva Kolledži vene keele lektoraadi esindajatega. Saadud teabe põhjal koostati vajaliku ja sobiva süsteemi kirjeldus, mis on esitatud selles peatükis.

Korpus RusLAPSED on keeleline korpus. Korpuse fookus on väga kitsas – Eestis elavate venekeelsete laste suuline ja kirjalik kõne. Korpuse peamised kasutajad hakkavad olema Tartu ülikooli tudengid.

See korpus on vajalik tekstide hõlpsaks ja kiireks otsimiseks sõna, sõnavormi ja lause ning teksti atribuutide ja metaandmete järgi. Sellise otsingu tegemiseks on vaja eelnevalt täita süsteem tekstidega, mis on varustatud otsingu jaoks vajalike lisaandmetega.

Tekstide kogumist ja sisu täitmist ei saa teha automaatselt, kuna iga teksti juurde lisatud andmed on väga olulised ja neid ei saa programmeerimise teel koguda. Seepärast kõnekorpusse tekstide kogumise ja lisamisega tegelevad inimesed, kes on loodava süsteemi registreeritud kasutajad. Väga olulised on ka tekstide allikad – need peaksid

olema Eestis elavad lapsed, kelle emakeel on vene keel. Kõik tekstid koostatakse ainult vene keeles, kuna peamine eesmärk on uurida vene keele arengut. Seepärast kasutajaliidese peamiseks keeleks on vene keel, kuid tulevikus plaanitakse liidesesse lisada ka eesti keele.

3.1 Infosüsteemi kasutajad

Korpuse RusLAPSED kasutajad jagatakse kolme rolli, igal rollil on süsteemis erinevad õigused: registreerimata kasutaja, registreeritud kasutaja ja administraator. Lisatud tekstide kontrollimine on vaid üks põhjus rollijaotuse vajaduseks korpuses. Materjalid korpuse sisu arendamise alustamiseks kogusid TÜ Narva Kolledži vene keele lektoraadi õppejõud ja üliõpilased. Õppejõud viivad läbi seminare ja üliõpilased kirjutavad nende materjalide põhjal uurimistöid, mistõttu peaksid õppejõud tegutsema administraatoritena ja õpilased sisu arendajatena.

- Registreerimata kasutaja

Kõige enam piiratud õigustega kasutaja (mitteautenditud ning mitteautoriseeritud kasutaja). Korpuse arendusetapis saab ta vaadata ainult üldist infosüsteemi teavet. Pärast korpuse töö alustamist lisandub ilma registreerimise ja autentimiseta kasutajale võimalus teksti sees otsida, mis on selle korpuse arendamise üks lõppeesmärkidest.

- Registreeritud kasutaja (sisu arendaja või toimetaja)

Teises rühmas on registreeritud kasutajad, kellel on kõik registreerimata kasutajate võimalused. Pärast autentimist saab kasutaja lisada ja muuta oma isikuandmeid, lisada uusi tekste ja need märgistada. Samuti saab autenditud kasutaja märgistada lisatud tekstides sõnavorme. Teksti ja metaandmete mistahes toimetamine on võimalik ainult nendes materjalides, mille kasutaja ise varem lisas.

Iga lisatud teksti peab kontrollima ja vajadusel kommenteerima teine autenditud kasutaja. Kommentaarid on nähtavad ainult teksti lisanud kasutajale. See on vajalik kõikide kirja- ja muude vigade kõrvaldamiseks.

- Infosüsteemi administraator

Korpuse administraatoril on mõlema ülalkirjeldatud kasutaja õigused. Uute kasutajate süsteemis registreerimine toimub ainult administraatori kaudu. Uue kasutaja lisamisel saab ta märkida inimese rolli korpuses: sisu arendaja või administraator. Samuti on tal võimalik infosüsteemi iga kasutaja blokeerida ja taastada.

Administraatoril on võimalik vaadata kasutajatega seotud teavet, samuti vaadata kasutaja viimase sisselogimise aega, lisatud tekstide ja sõnade arvu, vaadata kasutajale kontrollimiseks saadetud tekste ja edasi suunata/jagada need tekstid teiste projektis osalejate vahel. Tekstide ümbersuunamine on vajalik koormuse jaotamiseks süsteemi sisuarendajate (tudengite) vahel. Lisaks sellele saab administraator toimetada tekste ja metaandmeid ning teha märgistusi või muuta kõigi tekstide sõnu. Keelekorpuse uudiste- ja infoplokiga tegeleb samuti ainult administraator.

3.2 Infosüsteemi osad

Laste keelekorpus on veebileht, mille saab jagada mitmeks suureks loogiliseks mooduliks:

3.2.1 Sisselogimise, registreerimise ja kasutajate andmete haldamise moodul.

- Autoriseerimine

Keelekorpuse arendusetapis süsteemi peamised funktsioonid on suletud autoriseerimata kasutaja jaoks. Tekstide kuvamiseks ja lisamiseks peab sisse logima oma kasutajatunnuse ja parooliga.

- Isiklik konto

Isikuandmete täitmine või muutmine on võimalik ainult pärast autoriseerimist kasutaja isikliku kontoga. Selles moodulis on erinevate õigustega kasutajate jaoks erinevad võimalused. Süsteemi administraatorid saavad vaadata kõigi kasutajate andmeid. Süsteemi sisuarendajad saavad täies ulatuses näha ja muuta ainult oma isiklikke andmeid.

Selles moodulis kuvatakse ka tekstide loend, mis saadetakse teistele kasutajatele kontrollimiseks ja teiste arendajate isiklikud sõnumid, millele kasutajad võivad vastata samuti selles moodulis.

3.2.2 Teksti sisestamise moodul

Teksti saavad lisada ainult registreeritud ja autoriseeritud kasutajad. Ühe teksti koguja poolt (inimene kes kirjutab üles laste kõned) võib olla lisatud mitu teksti. Samuti üks laps saab olla mitmetes tekstides osalejana – see on väga vajalik info edaspidise teksti töötlemiseks ja analüüsimiseks. Uute materjalide (tekstide) lisamisel on võimalus valida mõned tekstide metaandmed (lapsed, kogujad jms) süstemis olemasolevatest (varem lisatud) andmetest. Teksti lisamine koosneb mitmest osast:

- Teksti metaandmete märgistus

Teksti metaandmetele lisatakse koguja või kogujad. Selle mooduli saab täita nii, et valitakse kas andmebaasis olemasolev koguja või lisatakse uus koguja. Samuti selles moodulis peab lisama osalenud lapse või lapsed ning mõned andmed iga lapse ja tema pere kohta. Nagu kogujate puhul, saab lapse andmebaasist valida või uue lapse lisada. Lisaks sellele on vaja lisada kõne üleskirjutamise andmed koha, aja ja salvestustingimuste kohta.

- Teksti lisamine

Teksti lisamise moodulis on vaja lisada kommentaar/ülesanne, olukorra kirjeldus selle kohta, millal laste kõne salvestati. Selle mooduli põhiosa moodustab teksti enda lisamine koos viidetega lastele ja kogujale. Iga salvestatud lause peab viitama sõnade autorile. Pärast kogu teksti sisestamist peab selle koos metaandmetega andmebaasi salvestama.

3.2.3 Sõnavormide märgistuse moodul

Sõnavormide märgistamine on võimalik ainult pärast teksti andmebaasi salvestamist. Märgistamisel ei võeta arvesse koguja sõnu ja mitmesuguseid teksti kommentaare. Sõnavormide märgistamine toimub ainult lausetes, mida räägivad lapsed. Pärast teksti salvestamist jagatakse laste laused programmiselt eraldi sõnadeks, mida pakutakse

kasutajatele märgistamiseks. Edaspidises teksti otsimisel kasutatakse samu sõnu. Märgistus tehakse kahes etapis:

- Sisaldab sõnavormide muutmise ja eemaldamise võimalusi

Kuna programmiliselt ei ole võimalik arvestada vene keele sõnavormide kõiki tunnuseid (mõned neist võivad koosneda mitmest sõnast), saab kõiki märgistamiseks pakutud sõnavorme muuta või eemaldada. Kui see sõnavorm on juba märgitud ja olemas andmebaasis, kuvatakse koos iga sõnavormiga selle kõneosa.

- Sõnavormi metaandmete märgistus

Märgistamisel tuleb kõigepealt märkida ära sõnavormi sõnaliik. Sõltuvalt sõnaliigist kuvatakse märgistuseks vajalikud tunnused. Kui see sõnavorm on andmebaasi juba varem lisatud ja sellel on näiteks mitu märgistusvalikut, pakutakse kasutajale valida antud sõnavormi jaoks sobiv märgistusvalik või luua uus. See funktsioon on vajalik, kuna samal sõnavormil võib olla erinev märgistus. Näidiseks võib tuua „много мам“ ja „мам, посмотри“. Sõnavorm „мам“ on sama, kuid sellel on täiesti erinevad atribuudid, erinevad metaandmed.

Samuti peaks selles moodulis olema võimalus märkida sõna algvorm, luua seos ühe rühma sõnavormide vahel. See on vajalik teksti otsimiseks sõnade järgi. Näiteks, otsida kõiki tekste sõnaga „мам“ ainsuses ja mitmuses. Sõnavormi märgistamine tehakse teksti otsimiseks metaandmete järgi.

3.2.4 Teksti otsimise moodul

Arendusetapis tekstiotsingu moodul on autoriseerimata kasutaja jaoks suletud. Pärast korpuse tarkvaraarenduse lõppu ja piisava korpuse materjali kogumist otsing avatakse kõigile kasutajatele. Moodul koosneb mitmest põhiosast:

- Otsing ja filtreerimise vorm

Teksti on võimalik otsida:

- vastavalt sõnavormile ja/või kõigile märgenditele, mis võivad andmebaasis sisalduda selle sõnavormi kohta
 - lapse ja kõigi lapse metaandmete järgi, mis on andmebaasis salvestatud lapse kohta
 - vastavalt laste kõne või teksti metaandmete salvestamise tingimustele, mis on andmebaasis salvestatud
- Otsingu tulemus/tekstide loend

Sellesse moodulisse esimest korda sattudes kuvatakse lühendatud kujul kõik tekstid, mis on andmebaasis kättesaadavad, see tähendab et näidatakse iga teksti kohta mõned read sellest. Tekstid on järjestatud lisamise kuupäeva järgi ja kuvatakse lehekülgede kaupa. Viimati lisatud tekstid kuvatakse esimesel lehel.

Otsingu ja filtrite rakendamisel kuvatakse loendis ainult need tekstid, mis vastavad valitud otsingu- ja filtreerimiskriteeriumitele.

- Teksti detailvaade

Näidatakse soovitud teksti tervikuna, lisades ära kõik teksti metaandmed. Samuti saab üksikasjaliku vaatamise ajal isiklikku seadmesse salvestada metaandmetega teksti MS Wordi dokumendi formaadis.

3.2.5 Projekti informatsiooni moodul

See moodul on saadaval kõigile süsteemi kasutajatele. Mooduli eesmärk on projektiga tutvumine ja see sisaldab üldist teavet projekti kohta. Selles moodulis kuvatakse mõned statistilised andmed, näiteks süsteemis olevate tekstide arv ja sõnavormide arv. Moodulite andmete muutmises osalevad ainult süsteemi administraatorid. Moodul jagatakse kolmeks:

- Korpuse uudised

Siin kajastub süsteemi muudatuste ja täienduste ajalugu.

- Informatsioon keelekorpusest

See osa sisaldab üldist teavet korpuse kohta: ajalugu, olulised sündmused, loomise idee jms. Võib-olla tulevikus ka teave projekti arendamist toetavate sponsorite kohta.

- Informatsioon arendajatest

See on kõige staatilisem osa, milles saab ainult teavet sisestada ja kustutada.

4 Olemasolevate lahenduste uuring

Esialgsete nõuete kogumine aitas koostada süsteemi vajalike komponentide nimekirja, samuti selgitada välja RusLAPSED keelekorpuse selge fookus. Mugava kasutajaliidese ja funktsionaalsusega infosüsteemi loomiseks otsustati uurida olemasolevaid lahendusi, et tuua välja tugevad ja nõrgad küljed ning võimalusel rakendada oma süsteemis parimaid lahendusi ja ideid.

Võrdluseks oli valitud mitu kõnekeele korpust, täpsemalt nende osi. Näiteks, on mõnedel korpustel üsna palju funktsioone või kuvatakse suures koguses infot, mida ei ole vaja projektile RusLAPSED. Seepärast mõned korpuste moodulid pakuvad huvi RusLAPSED projekti arendamiseks ja mõned mitte ning võrdluseks valiti mitte kõiki korpuste võimalusi.

4.1 Tallinna Ülikooli eesti vahekeele korpus (EVKK)

Esimene korpus – Tallinna Ülikooli eesti vahekeele korpus. „Tallinna Ülikooli eesti vahekeele korpus (EVKK) on eesti keele kui riigikeele (teise keele) ja võõrkeele õppijate kirjalike tekstide kogu. EVKK-s on rida alamkorpusi, kasutajaliides, mitmetasandiline annoteerimis- ja märgendussüsteem, statistikamoodul, tekstide automaatanalüüsi võimalused jm. Kombineerides erinevaid alamkorpusi, tekstilisi tunnuseid, vealiike ja metateavet õppija kohta, võimaldab korpuse kasutajaliides teostada mitmetasandilist otsingut.“ (EVKK veebileht)

Korpuse materjale saab kasutada teaduslikuks uurimistööks, magistri- ja doktoritööde kirjutamiseks, samuti õpetajate koolitamiseks (EVKK veebileht).

4.1.1 Süsteemi veebilehe sisu üldine kirjeldus

Veebilehel on kaks kasutajaliidese keelt – eesti ja inglise keel. Veebilehe menüü asub lehe vasakus servas ja sisaldab suurt hulka viiteid (Joonis 1). Samuti veebilehe vasakus servas on olemas kasutajanime ja parooliga sisselogimise vorm. Registreerimisvõimalust puudub.

The screenshot shows the 'Eesti vahekeele korpus' (Estonian Interlanguage Corpus) search page. The header has 'est eng' language toggles. A left sidebar contains a menu with links: 'Mis on EVKK', 'Korpuse kasutaja', 'Töötajad', 'Materjalid', 'Kasulikud lingid', 'Päring', 'Statistika', 'Morfosüntaktiline analüüs', 'Silbid, lemmad', 'Silbitaja', 'Pöördsonastik', 'Sõna- ja vormisagedus', 'Veataksnoomia', 'Kasutajanimi:', 'Parool:', and a 'Login' button. The main area is titled 'Päring' (Search) and contains several search criteria with dropdown menus: 'Sõna:' (text input), 'Korpus:' (dropdown with 'kõik' selected), 'Teksti keel:' (dropdown with 'eesti' selected), 'Elukoht:' (dropdown with 'pole oluline' selected), 'Vanus:' (dropdown with 'teadmata' selected), 'Emakeel:' (dropdown with 'teadmata' selected), 'Keele valdamise tase:' (dropdown with 'teadmata' selected), 'Abivahendid:' (dropdown with 'teadmata' selected), 'Märgendatud sõna:' (text input), 'Teksti tüüp:' (dropdown with 'teadmata' selected), 'Sotsiaalne taust:' (dropdown with 'pole oluline' selected), 'Sugu:' (dropdown with 'teadmata' selected), 'Kodune keel:' (dropdown with 'teadmata' selected), and 'Haridus:' (dropdown with 'teadmata' selected). At the bottom of the filters are an 'Otsi' (Search) button and a 'Dokumentide loend' (Document list) dropdown.

Andmed: EVKK veebilehe autori ekraanipilt.

Joonis 1. EVKK korpuse päringu vaade.

Veebilehel on kuvatud üldine teave projekti kohta ja kasulikud viited. Otsinguvorm sisaldab välja otsingusõna sisestamiseks ning lisafiltreid, mida saab näha Joonis 1.

Vajalike tekstide saamiseks filtris „Teksti keel” on vaja valida „vene“, filtris „Vanus” on vaja valida „kuni 18“, filtris „Emakeel” on vaja valida „vene“. Täiendavad filtrid saab valida oma soovi järgi. Nende filtrite lisamisega saame alla 18-aastaste venekeelsete laste vene keeles kirjutatud tekstid. Pärast „Otsi“ nupule vajutamist kuvatakse pealkirjade loend. Lehe ülaosas kuvatakse leitud tekstide koguarv. Kui klõpsata pealkirjal, avaneb aken, kus kuvatakse terve tekst ja teksti metaandmed. Andmete kuvamisviis on näidatud **Joonis 2**.

Andmed: EVKK veebilehe autori ekraanipilt.

Joonis 2 EVKK korpus. Teksti detailne vaade.

Sellel korpusel on tehtud sõnade märgistus, kuid otsitava sõna märgistuse valik pole selge ja seda pole kuskil kirjeldatud. Tekstisõnade märgistuse vaatamiseks on vaja vajutada viitele “morfo”, mis on **Joonis 2** tähistatud punasega.

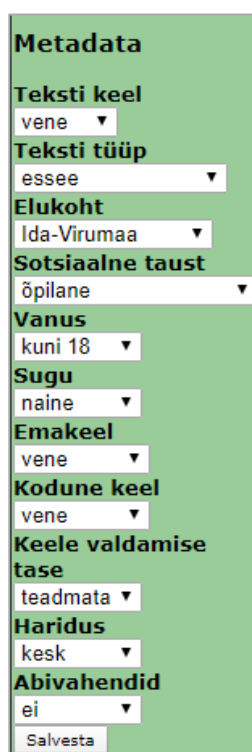
Sõnade märgistus on tehtud vaid eestikeelsetes tekstides, kuna teksti sõnadel on teatav metakeel, mida võib näha **Joonis 3**. Venekeelsetes tekstides õige märgistus puudub, arvatavasti selle pärast, et veebilehe peamine sihtmärk on uurida eesti keele kasutamist võõrkeelt kõnelevate inimeste poolt.

```
"<Teie>"
"sina" L0 P pers ps2 pl gen cap @P> #18->19
"<poolt>"
"poolt" L0 K post cap <gen> @ADVL #19->20
"<pakutud>"
"pakku=tud" L0 A pos sg nom cap partic <tud> @AN> #20->21
"<teema>"
"teema" L0 S com sg nom cap @SUBJ #21->22
"<paistab>"
"paist" Lb V main indic pres ps3 sg ps af cap <FinV> <Intr> @FMV #22->22
"paist" Lb V mod indic pres ps3 sg ps af cap <FinV> <Intr> @FCV #22->22
"<mulle>"
"mina" L1e P pers ps1 sg all cap @ADVL #23->22
"<huvitavana>"
"huvitav" Lna A pos sg es cap @ADVL #24->22
"<.>"
```

Andmed: EVKK veebilehe autori ekraanipilt.

Joonis 3 EVKK korpus. “Ametlik kiri“ teksti sõnade märgenduste vaade.

Teksti metaandmeid tähistatakse, valides ripploendist sobiva väärtuse. Metaandmete lisamise meetod on näidatud joonisel 4.



The image shows a vertical form titled "Metadata" with a green background. It contains several dropdown menus for selecting metadata values:

- Teksti keel**: vene
- Teksti tüüp**: essee
- Elukoht**: Ida-Virumaa
- Sotsiaalne taust**: õpilane
- Vanus**: kuni 18
- Sugu**: naine
- Emakeel**: vene
- Kodune keel**: vene
- Keele valdamise tase**: teadmata
- Haridus**: kesk
- Abivahendid**: ei

At the bottom of the form is a button labeled "Salvesta".

Andmed: EVKK veebilehe autori ekraanipilt.

Joonis 4. EVKK korpus. Metaandmete märgistuse vaade.

See korpus ei vasta enamikele süsteeminõuetele RusLAPSED, kuid see valiti ikkagi võrdluseks, kuna sellel on tekstifilter, mis võimaldab piirata otsingut sõnade autori vanuse ja teksti keele järgi. Sellisel viisil võib korpusest välja otsida alla 18-aastaste venekeelsete autorite tekste.

4.1.2 Korpuse väärtused

EVKK korpuses on võimalus vaadata kõiki saadaolevaid sõnu tähestiku järgi – vajutades sobivale sõnale, avaneb kõikide seda sõna sisaldavate tekstide loend.

Samuti sellelt veebilehelt saab alla laadida tekstide kasutamise loa vormi „Luba teksti kasutada“ ja vormi „Taustainfo“ teksti metaandmete mugavaks täitmiseks materjali salvestamisel. See on tekstikogujatele väga mugav funktsionaalsus, kuna neid dokumendivorme pole vaja kuskilt otsida ega ise koostada.

4.1.3 Korpuse puudused

Veebilehe peamine puudus on päringute töötlemise väga madal kiirus. Teine puudus on teabe puudumine sõnade märgistuse metakeele kohta. Samuti pole arusaadav miks teksti lisamine lihtkasutaja poolt pole võimalik, aga redigeerimine on lubatud. Võib-olla on see lihtsalt tarkvaraviga, käesoleva töö autori poolt üsna kriitiline, kuna iga kasutaja saab teksti muuta. Autori jaoks aga oli see väga kasulik, kuna autor sai näha, kuidas teksti lisamine ja selle metaandmete märgistamine on lahendatud, kuid selles moodulis polnud sõna märgistamise võimalust, seetõttu pole teada, kuidas seda tehti.

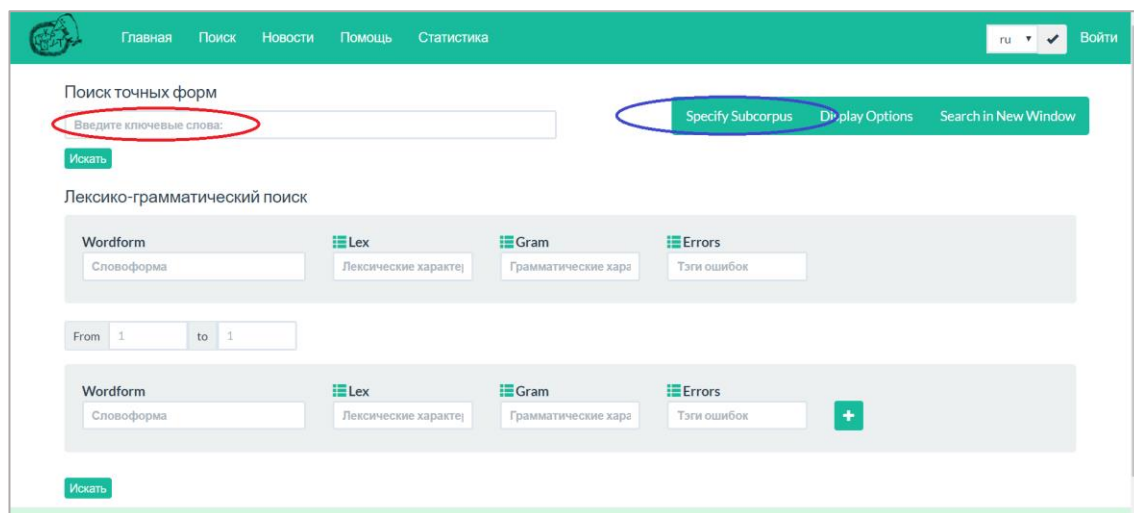
4.2 Corpus of Russian Student Texts (CoRST)

Teine korpus on vene õppetekstide korpus CoRST. See on mitme ülikooli tudengite vene keeles kirjutatud õppetekstide keelekorpus. See korpus on avatud kõigile kasutajatele ning on mõeldud keele- ja sotsiolingvistilisteks uurimiseks. Kogu korpuse maht on 3,1 miljonit sõna. (CoRST veebileht).

4.2.1 Veebilehe sisu üldine kirjeldus

Veebilehel on kaks kasutajaliidese keelt – vene ja inglise keel. Süsteem sisaldab väga lihtsat ja lühikest menüüd, mis asub lehe ülaosas, seal on ka vorm süsteemi sisenemiseks. Tekstide lisamine registreerimata kasutajale on võimatu, samuti on võimatu ennast korpuses registreerida. Veebilehe avalehel kuvatakse üldine teave projekti kohta. Samuti on seal uudiste leht, statistika ja otsingujuhend.

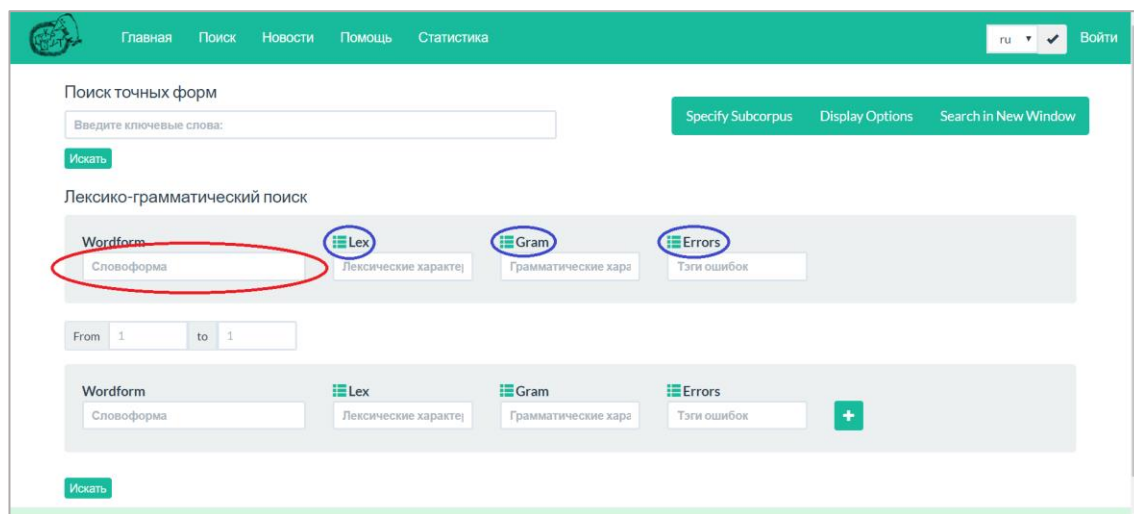
Korpuses saab otsida sõna täpse vormi või suvalise vormi järgi. Sõna täpse vormi otsimisel, mis on näidatud *Joonis 5*, peab sisestama sõna otsinguribale (tähistatud punasega) ja ära märkida kõik vajalikud metaandmed, mis asuvad paremal (tähistatud sinisega). Metaandmete filtreerimine on väga ulatuslik ja sellel on mitu valikut. Filtrite valimine on väga lihtne, see toimub valikuruutu kasutades. Filtreerimisvõimalust võib näha lõputöö lisas 1 joonisel 1.



Andmed: CoRST veebilehe autori ekraanipilt.

Joonis 5. CoRST korpus. Otsingu mooduli vaade 1.

Teist otsinguviisi kasutades, nimelt leksikalises ja grammatilises otsingumeetodis tuleb kesksesse otsinguaknasse sisestada nõutav sõnavorm, mis on demonstreeritud **Joonis 6** ja tähistatud punasega. Nupud filtreerimise avamiseks morfoloogiliste tunnuste järgi on märgitud sinisega.

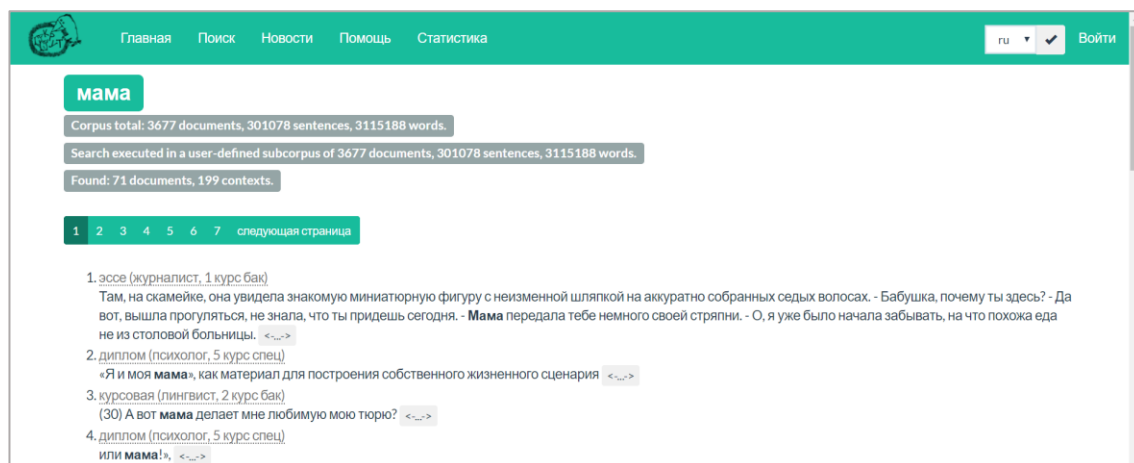


Andmed: CoRST veebilehe autori ekraanipilt.

Joonis 6. CoRST korpus. Otsingu mooduli vaade 2.

CoRST korpusel on võimalus otsida mitme sõnavormi järgi, mida saab lisada alumises otsingukastis asuva nupu „+“ abil.

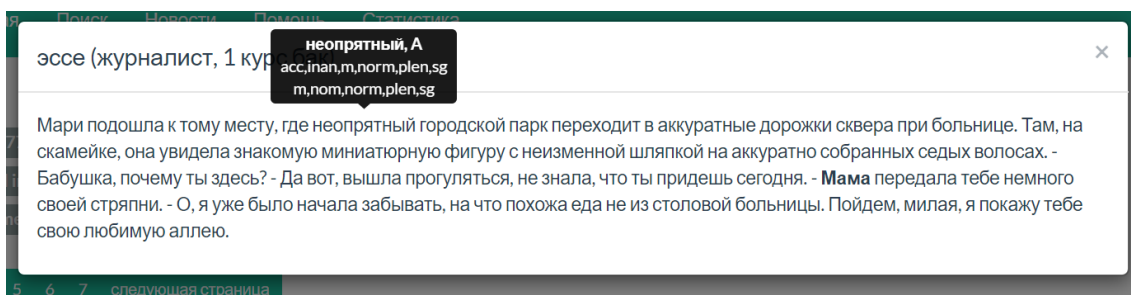
Otsingutulemid kuvatakse loendina. Loendi ülaosas on teave tulemuse kohta. **Joonis 7** on kuvatud näidises oli otsitud sõna “мама”, ning see kuvatakse lehe ülaosas.



Andmed: CoRST veebilehe autori ekraanipilt.

Joonis 7. CoRST korpus. Otsingu tulemus.

Vajadusel saab nimekirja iga elementi laiendada, kuid täisteksti ei kuvata. Kuvatakse ainult teksti fragment, milles on otsitud sõna esile tõstetud. Kui liigutada kursorit üksikute sõnade kohale, kuvatakse nende morfoloogiline märgistus. Sõna „неопрятный“ märgistuse näide on toodud **Joonis 8**.



Andmed: CoRST veebilehe ekraanipilt.

Joonis 8. CoRST korpus. Teksti osa vaade ühe sõna märgistuse kuvamisega.

Korpuse eeliseks on kasutusmugavus ning korpuse kasutamine ei vaja eriväljaõpet, pole vajadust teada sõnade märgistuse metakeelt selleks, et teha otsingut, kuid sõnade märgistust teksti lõigus on raske mõista, kuna see on kirjutatud metakeeles ja sellel puudub seletus.

CoRST korpus ei vasta süsteemi RusLAPSED nõuetele, kuid süsteemi otsingus on palju metainfo valiku võimalust ja üsna mugav kasutajaliides.

4.2.2 Korpuse väärtused

Menüüriba punktis „Помощь“ on kasutusjuhend, mis kirjeldab otsinguprotsessi süsteemis. Antud süsteem on küllaltki kasutajasõbralik, kuid raskuste tekkimisel võib igal ajal kasutusjuhendi poole pöörduda.

4.2.3 Korpuse puudused

Selle veebilehe peamine puudus on võimaluse puudumine kogu teksti vaadata, on võimalik vaadata vaid üksikuid lauseid, mis sisaldavad otsingusõna. Samuti süsteemil on olemas puudused kasutajaliidese tõlkimisel. Mõnel väljal ja nupunimel pole tõlget, liidese keel on segane, osa teavet vene keeles, osa inglise keeles. Kasutusjuhendil pole ka tõlget inglise keelde. Tõlkimise puudust võib näha kõigil ülaltoodud veebilehe ekraanipiltidel.

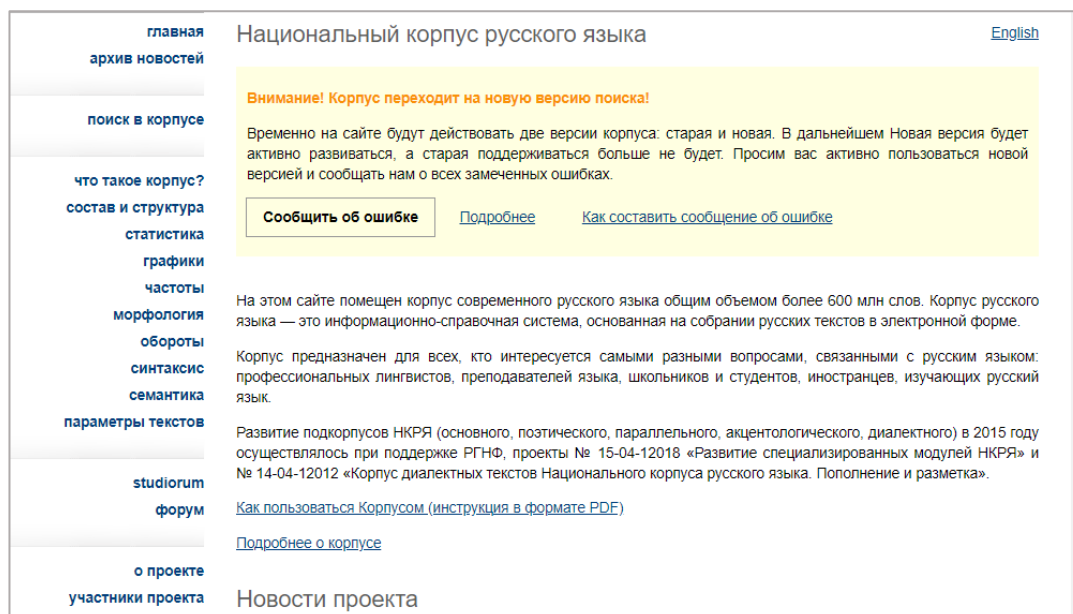
4.3 Vene keele rahvuskorpus (Национальный корпус русского языка)

Kõigist esitatud mahukaim korpus on vene keele rahvuskorpus. Süsteem on mõeldud kõigile, kes on huvitatud vene keelega seotud probleemidest ning sõnavara, keele grammatika ja keelemuutuste uurimiseks. Korpuse maht – 288 727 494 sõnakasutust ja 119 203 dokumenti. (VKRK-e veebileht)

4.3.1 Veebilehe sisu üldine kirjeldus

Veebilehel on võimalus kasutajaliidese keelt vahetada. Süsteemi peamine keel on vene keel, teine keel on inglise keel, kuid veebileht ei ole täielikult tõlgitud – mõni funktsioon on tõlgitud, kuid suurem osa ei ole. Avalik sisselogimise võimalus kasutajatunnusega ja parooliga puudub, samuti ei ole näha registreerimise võimalust.

Veebilehe menüü asub vasakus servas, sellel on palju viiteid, mis viitavad allikatele, mis sisaldavad erinevat informatsiooni. Veebilehe menüü on kuvatud *Joonis 9*.



Andmed: VKRK veebilehe autori ekraanipilt.

Joonis 9. VKRK korpus. Esimese lehekülje vaade menüüga.

Kui valida menüüst punkti „поиск в корпусе“, siis menüü muutub. Menüüloendis kuvatakse ülemineku avalehele või erinevatesse korpustesse, mille sees saab otsida.

Korpuse veebileht sisaldab üldist informatsiooni, uudiseid, korpuse kirjeldust, kõigi projektis osalejate nimekirja, korpuse põhjal tehtud väljaannete loendit ja suurt loetelu viidetega muudele tekstikorpustele.

Moodulis „участники проекта“ saab näha loendit, mis sisaldab isikuid, kes selle projekti väljatöötamisel osalesid nagu, näiteks, tarkvara arendajad ning sisemise märgistuse ja sõnavormide toimetajad. Projektis osalejate hulgas on ka Tartu Ülikooli esindajaid – Maria Borovikova (PhD, Tartu Ülikooli slaavi filoloogia osakonna teadur) ja Jelizaveta Fomina (PhD, Tartu Ülikooli slaavi filoloogia osakonna lõpetaja). (VKRK-e veebileht)

Korpus sisaldab asjalikku kasutusjuhendit, mis koosneb 84 leheküljest. Juhend avatakse PDF-formaadis, ning seda saab leida **Joonis 10** siniselt näidatud viite kaudu.

Materjale saab otsida sõna või lause järgi, teine võimalus on kasutada leksikaalset ja grammatilist otsingut.

Andmed: VKRK veebilehe autori ekraanipilt.

Joonis 10. VKRK korpus. Otsingu vaade 1.

Sõna või lause otsimisel peab sisestama soovitud väärtused otsinguribale, mis on **Joonis 10** märgitud punasega. Pärast päringu sisestamist tuleb vajutada nuppu „искать“.

Funktsiooni „задать подкорпус“ abil, mis on näidatud joonisel 10 roheliselt, saab otsinguteksti piirata erinevate metaandmetega. Teksti filtreerimise leht on väga suur ja sisaldab suurt metaandmete valikut (lisa 1 joonis 3).

Andmed: VKRK veebilehe autori ekraanipilt.

Joonis 11. VKRK korpus. Teksti žanri ja tüüpi otsingu filtrid.

Allpool olevas kirjelduses on näha vaid selle lehe osasid. Näiteks filtreerimislehel asuvad teksti žanri ja tüübi valimise moodulid (**Joonis 11**). See moodul võimaldab sisestada soovitud andmeid või valida neid loendist. Loendist valiku tegemiseks peab vajutama viitele „выбрать“, mille tulemusena avaneb mitme valikuga aken.

Leksiko-grammatiline otsing süsteemis toimub CoRST korpusega sarnaselt. Peab sisestama otsingukasti sõna soovitud vormi ja lisada kõik vajalikud andmed sõna morfoloogiliseks märgistamiseks. Märgistusandmeid saab sisestada vastavatesse akendesse metakeelte abil. Teine võimalus on valida need andmed valikvastustega loenditest, vajutades viitele „выбрать“. Otsingukasti täitmise näide on toodud **Joonis 12**, paigutuse valimise viit on märgitud punasega.

Andmed: VKRK veebilehe autori ekraanipilt.

Joonis 12. VKRK korpus. Otsingu vaade 2.

Samuti saab otsida erinevates projektikorpustes, lülitades need sisse menüüribal. Kuvatakse otsingutulem, mis on sarnane CoRST korpuse näitega. Tulemused kuvatakse loendina, mille ülaosas kuvatakse üldine informatsioon tulemuse kohta.

VKRK's samuti pole võimalik vaadata kogu teksti, saadaval on ainult väike lõik. Sõnale klõpsamisel kuvatakse morfoloogiline märgistus. Sõna märgistusega teksti kuvamise näide on toodud **Joonis 13**.

Результаты поиска в основном корпусе

Объем всего корпуса: 119 203 документа, 288 727 494 слова.

мама

Найдено 1 документ, 2 вхождения.

[Распределение по годам](#) [Статистика](#)

Поискать в других корпусах: [акцентологическом](#), [газетном](#), [диалектном](#), [мультимедийном](#), [обучающем](#), [параллельном](#), [поэтическом](#), [синтаксическом](#), [устном](#).

Страницы: 1

1. коллективный. Форум: Рецензия на фильм «Поймай меня, если сможешь» (2006-2011) [омонимия снята] [Все примеры](#)

[AleksandrNobody] Еще в школе с помощью обмана выдавал себя за учителя французского языка и проводил уроки целые месяцы. По началу для него это было всего лишь розыгрыш и дурачество. Но после того, как разведятся его родители, все меняется. Он убегает из дома и начинает жизнь, окутанную во лжи, но с определенной целью — вернуть счастье в их дом, и чтобы **мама** и папа снова были вместе. Для этого он начинает зарабатывать деньги, кроме того, он начал м...
мошенничеством и преступлением...
ум сразу приходят такие великие ф...
Райана» и др. [коллективный. Форум: Ре...

начинает	
Лемма	начинать
Грамматика	V,3p.act.indic.ipf.praes.sg.tran
Семантика основная	aux:phase d:pref
Доп. признаки	num:red

Страницы: 1

Поискать в других корпусах: [акцентологическом](#), [газетном](#), [диалектном](#), [мультимедийном](#), [обучающем](#), [параллельном](#), [поэтическом](#), [синтаксическом](#), [устном](#).

[Сообщить об ошибке](#)

Скачать несколько первых результатов выдачи в формате [Excel](#), [OpenOffice Calc](#), [XML](#).

Andmed: VKRK veebilehe autori ekraanipilt.

Joonis 13. VKRK korpus. Teksti osa vaade ühe sõna märgendamise kuvamisega.

4.3.2 Korpus väärtused

Morfoloogiamoodul kirjeldab kogu sõnade morfoloogilist märgistust ja sisaldab kirjeldusi metakeeles. Metakeele mõistmiseks on olemas vastav juhend.

Veebileht sisaldab asjalikke juhiseid PDF-formaadis. Tänu sellele saab absoluutselt iga kasutaja selle keelekorpusse töötamisega hakkama. Otsinguleht sisaldab näpunäiteid. Kui klõpsata mõne välja kõrval asuvale küsimärgile, siis avaneb näpunäidetega aken. Samuti on selles süsteemis, erinevalt kõikidest käesolevas töös varem esitatud süsteemidest, on olemas võimalus graafikuid tekitada. See funktsioon pakkus huvi ka RusLAPSED korpusse TÜ Narva kolledži lektoraadile esindajatele, seetõttu võib see funktsionaalsus olla näitena sellise idee elluviimisel.

Graafiku näide sõna „мама“ kasutamisest tekstides aastatel 1997 kuni 2017 on esitatud lisa 1 joonisel 2.

VK RK korpusel on veel üks huvitav omadus, mida on kasulik mainida – võimalus salvestada osa otsingutulemusest Exceli, OpenOffice Calc, XML formaatides.

4.3.3 Korpuse puudused

Praegu pole see veebileht lõplikult valmis. Keele vahetamisel inglise keelele kõik menüü viited viitavad lehele „other corpora“, millel puudub tõlge. Venekeelsele liidesele tagasi minna pole võimalik, sest lülitusvõimalus puudub. Lehe vaadet inglise keelele üleminekul võib näha *Joonis 14*.



Andmed: VKRK veebilehe autori ekraanipilt.

Joonis 14. VKRK korpus. Inglise keelne veebilehe vaade.

4.4 Korpuste võrdlemine

Kõik uuritud korpused on omal moel head, kuid mitte ükski neist ei ole suunatud laste kõne kogumisele, kuigi esimene neist – EVKK – sisaldab laste tekste, kuid ei sisalda spontaanset kõnet ja venekeelsete tekstide korrektset märgistamist. Viimasel korpusel VKRK’l on võimalus määrata sõnade autori sünniaasta, kuid kui sisestada alla 18-aastaste laste sünniaastad, annab otsing nulltulemuse. Mõningate võimaluste saadavuse võrdlus korpuse järgi on toodud tabelis 1.

Tabel 1 Korpuste võimaluste võrdlus.

Korpuse informatsioon	EVKK	CoRST	VKRK	RusLAPSED
Korpuse üldinformatsioon	+	+	+	+
Korpuse uudised	-	+	+	+
Korpuse maht	-	+	+	+
Kasutusjuhend	-	+	+	-
Kasutajaliidese keele valik	+	+	+	+
Statistika	+	+	+	-
Sõna märgistamise võimalused	+	+	+	+
Tekstide autorite metaandmed	+	+	+	+
Laste kõne tekstid	+	-	-	+
Spontaanne kõne	-	-	+	+
Otsing sõna alusel	+	+	+	+
Otsing sõnavormi alusel	+	+	+	+
Otsing lause alusel	-	-	+	+
Graafiku modelleerimine	-	-	+	+
Grandid ja fondide toetus	+	Info puudub	+	-

Allikas: Koostatud autori poolt

Teave toetuste kohta lisati teadlikult sellesse tabelisse, sest vaatamata sellele, et antud töös väljatöötataval projektil pole veel ühtegi toetust, tulevikus võib RusLAPSED sellele siiski loota.

5 Eeluurimuse kokkuvõte

Tuginedes uuritud laste kõnekorpustele võib öelda, et ühelgi neist pole soovitud funktsioone täielikus ulatuses. Mõne nõude rakendamist võib leida muudes tekstikorpustes, kuid põhimõtteliselt on see keskmise kasutaja jaoks ebamugavalt ellu viidud. Korpuste kasutamiseks on vaja teha lisauurimistööd, näiteks detailotsingu tegemiseks on vajalik juhiste üksikasjalik uurimine ja teatud käskude või märkide meelde jätmine. Samuti on paljudes süsteemides otsing üsna piiratud ja sellel pole piisavalt valikuvõimalusi.

Kõik korpused erinevad kasutajaliidese funktsionaalsuse ja leitud teksti vaatamise ulatuse järgi. Mõned korpused ei võimalda kogu teksti näha, mõnel juhul peab teksti lugemiseks teksti alla laadima (kuid need korpused ei ole selles töös käsitletud).

Pärast olemasolevate lahenduste uurimist võib kindlalt öelda, et laste kõne korpuse RusLAPSED arendamine on õigustatud, kuna autor ei ole nõutava funktsionaalsusega lahendust leidnud.

Korpusele RusLAPSED plaanitakse luua lihtne kasutajaliides, lihtsad päringuvormid, millega saab igaüks hakkama ilma spetsiaalse ettevalmistuseta. Samuti tulevases süsteemis teksti- ja sõnavormidele lisatakse palju lisaandmeid ja kõigi nende andmete jaoks on võimalik teha päringuid vastavalt uurija vajadustele.

Ühe lõputöö raames ei ole võimalik täita kõiki lastekõne korpusele RusLAPSED esitatavaid nõudeid, seepärast projekt on algusest peale plaanitud mitmele aastale, kus selle erinevaid osi arendavad erinevad üliõpilased. Antud töö autoril oli au seda projekti alustada ja sellele aluse luua.

II. ARENDATAVA SÜSTEEMI ANALÜÜS

1. Projekti haldamise vahendid

Projekti väljatöötamiseks on kasutatud mõned tarkvaralised abivahendid:

- Azure DevOps – kogu arendusprotsessi süsteem, mis sisaldab projektijuhtimise ning koodi haldamise, kompileerimise, testimise ja tarne moodulit. See programm võimaldab jälgida projekti seisundit ja iga üksikut moodulit. Seda süsteemi oli kasutatud ka mugavuse huvides ülesannete osadeks jagamise jaoks ja vigade tekkimisel viidi nende kirjeldused kohe sellesse süsteemi sisse. Korpuse RusLAPSED töölaua vaade on lisas number 2 joonisel 1.
- Git – versioonikontrollisüsteem, mis on samuti integreeritud Azure DevOps sisse. Võimaldab luua üksteisest sõltumatu programmikoodi harud, ning annab võimalusi neid ühendada ja kustutada (Git'i veebileht).
- SourceTree – versioonikontrolli klient. See programm oli kasutatud projekti versioonide haldamiseks. Iga ülesande jaoks oli tööd alguses loodud uus projektiharu, milles autor koodi kirjutas. Seega vigade korral oli alati võimalik projekti eelmisele versioonile tagasi viia. Selle rakenduse töölauda, koos RusLAPSED projekti koodiharudega võib näha lisa 2 joonisel 2. Pärast ülesande edukat lõpetamist haru koodiga ühendati projekti peaharuga, kasutades arendusprotsessi süsteemi Azure DevOps.
- Microsoft Visual Studio – integreeritud keskkond tarkvara arendamiseks. d
- Microsoft SQL Server Management Studio – andmebaasihaldussüsteem. Kasutatakse SQL Serveri andmebaasidele ja komponentidele juurdepääsu, konfigureerimise, haldamise ja arendamise jaoks (SQL, Microsofti veebileht).

2. Nõuded

Peatükkis „Infosüsteemi ootused“ on kirjeldatud korpuse visiooni. Nagu varem oli mainitud, ühe lõputöö raames ei olnud võimalik täita kõiki süsteemile esitatud nõudeid. Sellega seoses otsustati välja tuua see projekti funktsionaalsus, mida on otstarbekas välja töötada selle esimeses etapis, et luua aluse tulevase kõnekorpusse RusLAPSED jaoks.

Nõuded olid kogutud peamiselt kohtumistel kahe TÜ Narva Kolledži esindajaga Olga Burdakova ja Liudmila Nikolaevaga, samuti ülikooli tudengitega, kes kirjutavad oma

lõputöid morfoloogiliste märgistuse väljatöötamise kohta lastekõne korpusele RusLAPSED. Lisaks isiklikele kohtumistele koguti nõudeid ka kirjavahetuse ja videokonverentside teel.

Intervjuueerimismeetodi kaudu tehti kindlaks põhiosa funktsionaalsetest nõuetest ning osa neist sõnastati ka olemasolevate lahenduste võrdleva analüüsi tulemuse alusel, mis on toodud antud lõputöö esimeses peatükis.

Mittefunktsionaalsed nõuded vene lektoraadi poolt polnud esitatud. Seetõttu kõik mittefunktsionaalsed nõuded oli koostatud autori poolt lähtuvalt parimatest praktikatest, mida autor on omandanud ettevõtte praktika ja töö ajal.

2.1 Funktsionaalsed nõuded

1. Süsteemis peab olema kolm erinevat kasutajarolli, kellel on erinevad õigused: tavakasutaja (registreerimata), toimetaja ja administraator.
2. Tavakasutaja näeb ainult korpuse avalehte ja sisselogimislehte.
3. Süsteemis peab olema kasutajatunnuse ja parooliga autentimisviis.
4. Süsteemis peab olema parooli taastamise võimalus.
5. Kasutajatunnuseks peab olema e-posti aadress.
6. Kasutajad ei saa ise ennast registreerida.
7. Süsteemil ei ole piiranguid registreeritud kasutajate arvule.
8. Toimetajal on kõik tavakasutaja õigused ning lisaks neile saab ta pärast autoriseerimist ja rolli tuvastamist:
 - vaadata ja muuta oma andmeid;
 - lisada tekstid süsteemi;
 - luua tekstide ja sõnavormide märgistust;
 - kasutada teksti otsingut süsteemis.
9. Administraatoril on kõik toimetaja õigused ja lisaks nendele õigus:
 - registreerida süsteemis uued kasutajad;
 - vaadata andmeid registreeritud kasutajate kohta;
 - vaadata iga registreeritud kasutaja kohta viimase sisselogimise aega;
 - lisada, muuta ja kustutada informatsiooni korpuse kohta.
10. Süsteem peab sisaldama korpuse uudiseid, üldist infot korpuse kohta ja infot projekti arendajate kohta.

11. Uudiseid ja infot korpuse ja arendajate kohta peaks olema võimalik lisada, muuta ja kustutada.
12. Korpuses peavad olema kuvatud statistilised andmed süsteemis olevate tekstide ja sõnavormide arvu kohta.
13. Registreeritud kasutajal peab olema võimalus vaadata ja muuta oma isiklikud andmed.
14. Peab olema võimalus lisada uusi tekste süsteemi.
15. Tekstide lisamine peab olema lihtne ja arusaadav kasutaja jaoks.
16. Peab olema võimalus lisada metainformatsioon iga sisestatud teksti kohta.
17. Kogu metainformatsioon peab olema sisestatav eelnevalt ettevalmistatud valikute abil või avatud teksti väljade sisestamise abil.
18. Teksti metaandmete moodul peab sisaldama võimalust lisada kogu informatsiooni, mis on kuvatud lisas 3 tabelis 1. Samuti peab olema võimalus lisada kõik žanrid, mis on kuvatud lisas 3 tabelis 2.
19. Igas tekstis peab olema võimalus lisada kommentaar laste kõne üleskirjutamise tingimuste kohta.
20. Iga teksti sõnal peab olema võimalus metaandmete märgistamiseks.
21. Süsteemis peab olema võimalus otsida sõnade, lausete ja metaandmete alusel. Kõik tingimused on toodud lisas 3 tabelis 3.
22. Teksti otsing peab olema tehtud ainult laste poolt kasutatud sõnade ja lausete alusel. Kogujate enda poolt kirjutatud tekste ei tohi teksti otsingu ajal arvestada.
23. Kommentaare, mis on laste kõne tekstide sees, ei tohi arvestada teksti otsingu ajal, selleks on vaja kasutada sisestatud tekstide vormistusreegleid, mis on toodud lisas 3 tabelis 4.
24. Otsitud tekstid peavad olema kuvatud nimekirjana koos teksti osaga.
25. Süsteemis peab olema kogu teksti vaatamise võimalus koos kogu märgistatud metainformatsiooniga.

2.2 Mittefunktsionaalsed nõuded

1. Andmebaasi varukoopia tegemine. Andmete kaotamise vältimiseks peab tegema andmebaasi varukoopiat üks kord päevas.
2. Kasutaja parooli hoidmine andmebaasis peab olema krüpteeritud kujul. Enne parooli salvestamist paroolile peab olema lisatud „sool“ (tähtede kombinatsioon, mis lisatakse programmi koodis ja pole nähtav ei andmebaasis ega kliendi

rakenduses) ja seejärel peab olema krüpteeritud kaasaegse krüpteerimisalgoritmiga.

3. Piirangute kehtestamine parooli muutmisel. Parool peab sisaldama vähemalt kaheksat tähemarki, sealhulgas suurtähed, väiketähed ja numbrid.
4. Süsteemi skaleeritavuse võimalus. Infosüsteemi kood peaks olema lihtsasti laiendatav, kuna eeldatakse, et projekti kallal hakkavad töötama erinevad arendajad.

3. Kasutusloode koostamine ja prototüüpimine

3.1 Veebilehe päis ja menüü

Kasutuslugu 1. Veebilehe päis ja menüü

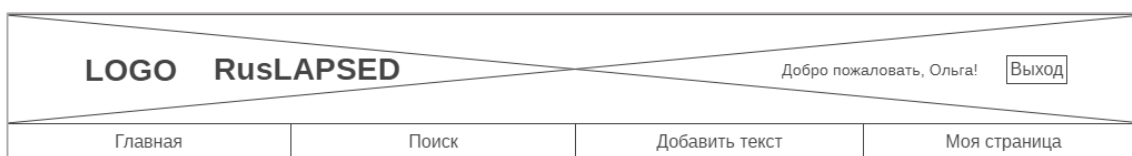
Peakasutajad: Registreerimata kasutaja, toimetaja ja administraator.

Eeltingimused: Registreerimata kasutaja jaoks eeltingimused puuduvad, toimetaja ja administraatori jaoks eeltingimuseks on autoriseerimine.

Kasutuslugu:

Registreerimata kasutaja võib näha ühe menüüpunkti „Главная“, veebilehe päist ja sisselogimise nuppu. Selle nupu vajutamisel avatakse süsteemi sisselogimise vorm. Toimetaja ja administraator võivad pärast autoriseerimist näha kõike menüüpunkte.

Kasutajaliidese eskiis:



Andmed: Autori joonis.

Joonis 15. *RusLAPSED korpuse päis ja menüü.*

3.2 Avaleht

Kasutuslugu 2. Avaleht

Peakasutajad: Kõik kasutajad, administraator.

Eeltingimused: Vaatamiseks puuduvad, muutmiseks administraatori autentimine.

Kasutuslugu:

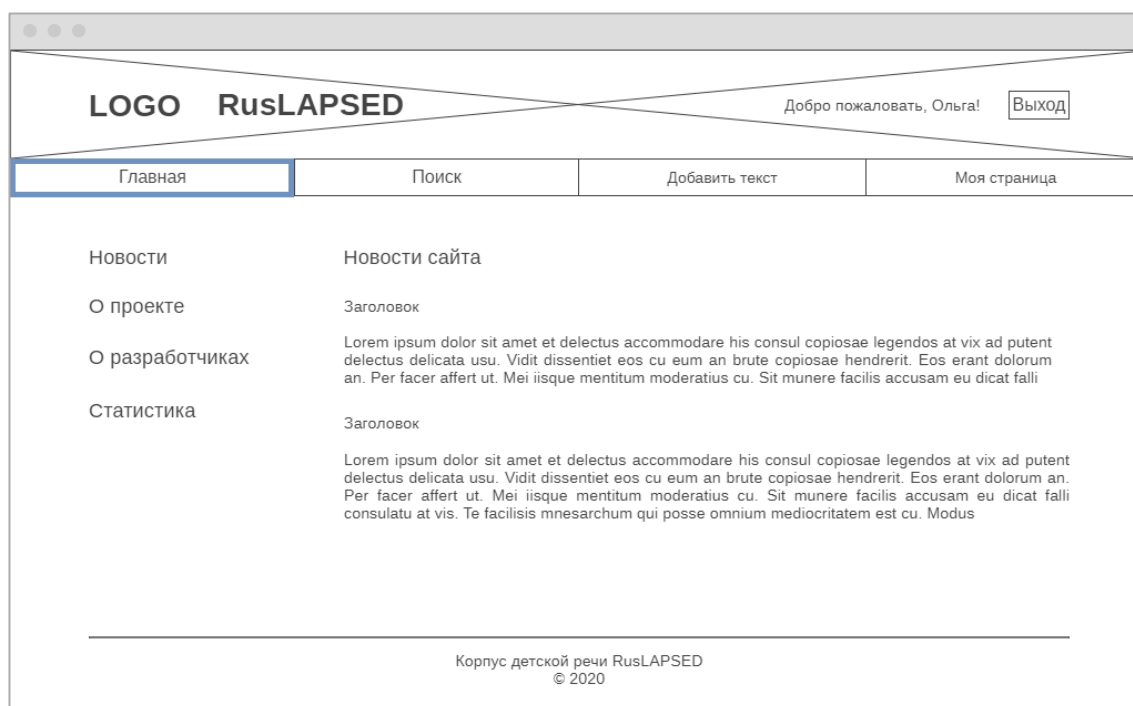
Avaleht või „Главная“ menüüpunkt. Süsteemi veebilehe URL’iga pöördudes automaatselt avatakse avaleht, mis sisaldab keele korpuse uudiseid. Info mooduli muutmise võimalus on ette nähtud vasakpoolses menüüs. Selles vertikaalses menüüs asuvad alalehtede viited: „Новости“, „О проекте“, „О разработчиках“, „Статистика“. Konkreetse menüü vajutamisel näidatakse vastav süsteemi moodul infoga.

Lehtede „Новости“, „О проекте“, „О разработчиках“ sisu muutmiseks saab tegeleda ainult autoriseeritud kasutaja rolliga „Administraator“. Igal ülesloetud lehel on olemas nupud:

- „Добавить новый блок“, mille vajutamisel avatakse vorm uue info lisamiseks. Selles vaates on näidatud nupp andmebaasi salvestamiseks ja nupp tagasi minekuks ilma muutmiste salvestamist.
- „Редактировать“, mille vajutamisel avatakse vorm teksti muutmiseks. Üldise info muutmise vaates asuvad andmebaasi salvestamise nupp „Сохранить“ ning tagasi mineku nupp „Назад“, mis võimaldab minna eelnevale lehele ilma muutmise salvestamiseta. Nupu „Назад“ vajutamisel ilmub teade, et kõik muudetud andmed ei salvestata andmebaasi.
- Nupp „Удалить“, mille abil saab kustutada nupu kõrval asuvat teksti. Nupu vajutamisel ilmub aken tegevuse kinnituseks.

„Статистика“ lehel on kuvatud statistilised andmed: süsteemis olemasolevate tekstide ja sõnakasutuste arv.

Kasutajaliidese eskiis:



Andmed: Autori joonis.

Joonis 16. *RusLAPSED* avalehe vaade.

3.3 Sisselogimine

Kasutusharu 3. *Korpusesse sisselogimine*

Peakasutajad: Registreeritud kasutajad.

Eeltingimused: Isikliku profiili olemasolu süsteemis.

Kasutusharu:

Veebilehte saavad näha kõik kasutajad, kuid sisselogimise võimalused peavad olema tagatud ainult süsteemis registreeritud kasutajatel.

Lehel on kuvatud väljad kasutajatunnuse ja parooli sisestamiseks ning nupp millega toimub süsteemi sisselogimine.

Kasutajatunnusega loetakse e-posti aadress. Juhul, kui sisestatakse vale kasutajatunnus või parool, näidatakse vastav veateade.

Sellel lehel on samuti võimalik parooli taastada. Selle toimingu tegemiseks peate sisestama välja „Признак пользователя“ kasutajatunnuse ja vajutama viitele "Забыли пароль?".

Pärast viite vajutamist kontrollitakse, kas süsteemis on olemas kasutaja sellise aadressiga. Kui kasutaja on olemas, siis saadetakse tema e-posti aadressile uus parool. Kui määratud e-posti aadressiga kasutaja süsteemis puudub, siis kuvatakse vastav teade.

Kasutajaliidese eskiis:

RusLAPSED
LOGO

Введите признак пользователя и пароль

Войти

[Забыли пароль?](#)

Корпус детской речи RusLAPSED
© 2020

Andmed: Autori joonis.

Joonis 17. RusLAPSED sisse logimise vaade.

3.4 Isiklik konto

Kasutuslugu 4. Isikliku konto erinevates rollides

Peakasutajad: Toimetaja, administraator.

Eeltingimused: Kasutaja autoriseerimine süsteemis.

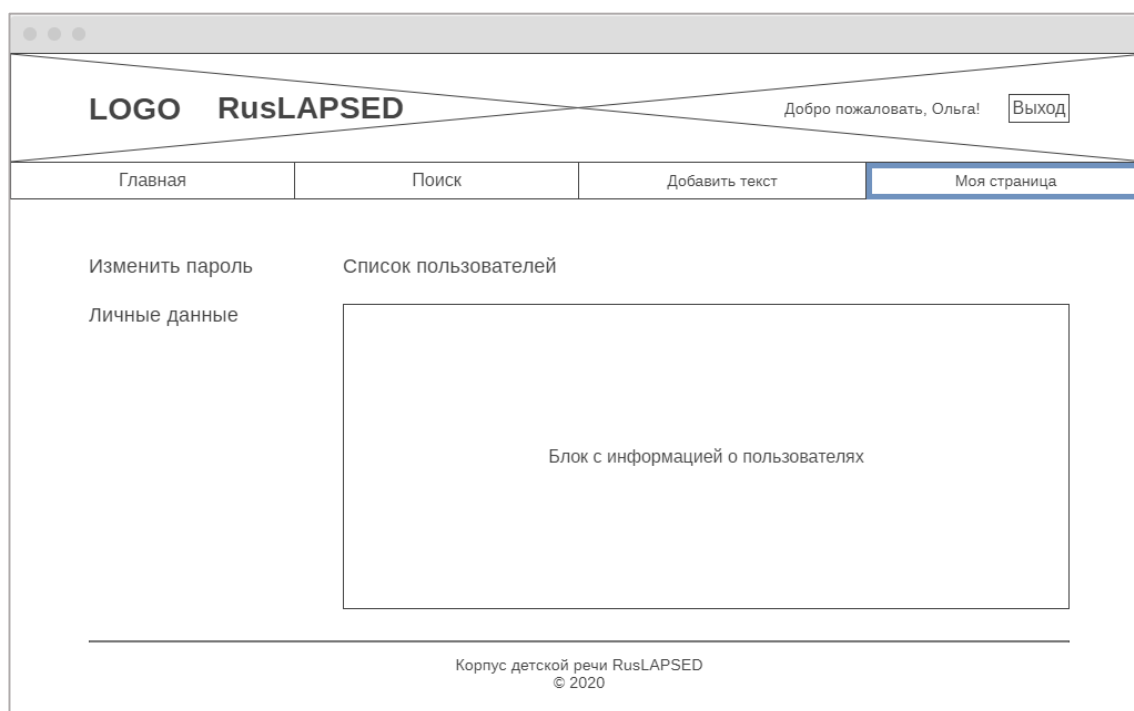
Kasutuslugu:

Isikliku teabe kuvamiseks või muutmiseks peab kasutaja vajutama menüüpunktile „Моя страница“. Pärast vajutamist kuvatakse kasutaja isiklik leht, kus asuvad viited parooli muutmise (Kasutuslugu 4.1) ja isikuandmete muutmise moodulile (Kasutuslugu 4.2).

Kasutaja isikliku lehe põhivaade on erinev toimetaja ja administraatori rollides. Praegu toimetaja rollis sellel lehel ei ole midagi kuvatud. Administraatori rollis sellel lehel on järgmised võimalused:

- Näha arendajate nimekirja koos isikuandmetega.
- Näha viimase sisselogimise aega ja kasutaja olekut (blokeeritud/aktiivne).
- Minna edasi kasutaja täieliku isikliku info kuvamise moodulile.
- Lisada, blokeerida ja taastada kasutajaid.
- Valida kasutaja roll uue kasutaja registreerimisel.

Kasutajaliidese eskiis:



Andmed: Autori joonis.

Joonis 18. RusLAPSED isikliku lehe vaade.

Kasutuslugu 4.1. Parooli muutmise moodul

Peakasutajad: Toimetajad, administraator.

Eeltingimused: Kasutaja autoriseerimine süsteemis.

Kasutuslugu:

Parooli muutmiseks peab kasutaja sisestama oma praegust ja uut parooli. Uus parool on vaja sisestada kahte välja. Pärast vastavate väljade täitmist on vaja vajutada nuppu „Сохранить“. Kui kõik väljad on õigesti täidetud, siis uus parool salvestatakse andmebaasi. Kui sisestatud paroolides on viga, siis kuvatakse veateade. Uus parool peab sisaldama vähemalt kaheksat tähemärki, vähemalt ühte suur- ja ühte väiketähte ning vähemalt ühte numbrit.

Kasutuslugu 4.2. Isiklike andmete muutmise moodul

Peakasutajad: Toimetajad, administraator.

Eeltingimused: Kasutaja autoriseerimine süsteemis.

Kasutuslugu:

Isikuandmete muutmise moodulis saab kasutaja märkida oma eesnime, perekonnanime ja e-posti aadressi. Pärast andmete muutmist tuleb vajutada nuppu „Сохранить“. Pärast nupu vajutamist kontrollitakse e-posti aadressi ning vea korral kuvatakse vastav teade.

3.5 Teksti lisamine

Kasutuslugu 5. Teksti lisamine

Peakasutajad: Toimetajad, administraator.

Eeltingimused: Kasutaja autoriseerimine süsteemis.

Kasutuslugu:

Uue teksti lisamine asub eraldi lehel, mida saavad näha ainult autoriseeritud kasutajad. Teksti lisamise moodul koosneb mitmest moodulist. Esimeses moodulis, mida kirjeldatakse kasutusloos 5.1, lisatakse koguja. Teine moodul sisaldab infot, mis on kirjeldatud kasutusloos 5.2 (lapse lisamise moodul). Kolmandas moodulis (kasutusloos 5.3) lisatakse informatsioon teksti kohta.

Järgmisena on toodud kaks tekstivälja teksti lisamiseks. Esimeses väljas on vaja kirjutada teksti koguja kommentaar teksti salvestamise tingimuste kohta. Teisele väljale on vaja sisestada tekst ise. Pärast teksti salvestamist on vaja märgistada sõnad nagu kirjeldatud kasutusloos 5.4.

Kasutajaliidese eskiis:

The wireframe shows a web application interface for adding text. At the top is a header bar with a logo on the left, the text 'RusLAPSED' in the center, and a user greeting 'Добро пожаловать, Ольга!' with a 'Выход' (Logout) button on the right. Below the header is a navigation bar with four links: 'Главная' (Home), 'Поиск' (Search), 'Добавить текст' (Add text), and 'Моя страница' (My page). The 'Добавить текст' link is highlighted with a blue border. The main content area contains several form blocks: 'Блок добавления собирателей' (Collector addition block), 'Блок добавления детей' (Child addition block), and 'Дополнительная мета информация' (Additional meta-information block). Below these is a section titled 'Блок добавления текста' (Text addition block) which includes a sub-label 'Комментарий об условиях записи текста' (Comment on recording conditions) and a text input field. Underneath is another label 'Окно для ввода текста' (Text input window) followed by a large text area. At the bottom right of the form is a 'Сохранить' (Save) button. The footer of the page contains the text 'Корпус детской речи RusLAPSED © 2020'.

Andmed: Autori joonis.

Joonis 19. RusLAPSED teksti lisamise lehe vaade.

Kasutuslugu 5.1. Kogujate lisamine

Peakasutajad: Toimetajad, administraator.

Eeltingimused: Kasutaja autoriseerimine süsteemis.

Kasutuslugu:

Selles moodulis saab teksti sisestaja valida teksti koguja nime olemasolevate teksti kogujate nimekirjast. Kogujate nimekiri kuvatakse ripploendina. On olemas ka nime järgi koguja otsimise võimalus. Otsingu jaoks on vaja sisestada välja otsitava nime osa. Teine võimalus koguja lisamiseks on sisestada koguja nimi tekstivälja ja vajutada nuppu „Добавить“. Kui andmebaasis on olemas sama nimega koguja, siis kuvatakse selle kohta teade. Sellel juhul kasutaja peab andmebaasist valima olemasolevat kogujat või sisestada uue koguja nimi.

Pärast koguja lisamist kuvatakse tema nimi teksti lisamise põhilehel. Lisatud kogujate arv pole piiratud. Samuti on võimalik eemaldada kogujat nimekirjast. Kui vajutada kustutamispupule, siis kuvatakse kinnitusaken.

Kasutuslugu 5.2. Laste lisamine

Peakasutajad: Toimetajad, administraator.

Eeltingimused: Kasutaja autoriseerimine süsteemis.

Kasutuslugu:

Lapse lisamise moodulis on samuti olemas kaks võimalust, nagu koguja lisamiselgi. Esimene võimalus – valida last nimekirjast, kus on kuvatud andmebaasis olemasolevad lapsed. Lapse valimisel võib kasutada otsingut nime või koodi järgi.

Igal lapsel on olemas isiklik kood süsteemis. See kood on üles ehitatud lapse sisestanud kasutaja ja lapse koodidest. Kasutaja kood on tema järjekorranumber süsteemis, lapse kood on lapse järjekorranumber, mille määramisel võetakse arvesse ainult praeguse kasutaja poolt lisatud lapsi. Seega, sisestades oma koodi otsinguväljale, saab kasutaja

nimekirja ainult nendest lastest, kelle on ta ise süsteemi lisanud. Teiste kasutajate poolt lisatud lapsi ei kuvata.

Lapse valimisel andmebaasist kuvatakse tema isikuandmeid, mida ei saa muuta (nimi, kood, sugu, sünniaeg, kaksikute olemasolu). Lisaks nendele andmetele on olemas moodul teabe sisestamiseks perekonna ja õppetingimuste kohta. Kuna need andmed võivad muutuda, siis need tuleb täita iga kord uuesti, kui lisatakse uus tekst sama lapse kohta.

Teine võimalus lapse lisamiseks on täita kõik väljad, mis tähendab uue lapse lisamist andmebaasi. Sel juhul on vaja täita ka isiklik teave lapse kohta: nimi, sugu, sünniaeg, kaksikute olemasolu. Kood antakse lapsele teksti salvestamise ajal programmiselt. Väli lapse nimega on kohustuslik täitmiseks. Samuti on olemas piirang lapse nimele – sama kasutaja ei saa lisada kahte sama nimega last, nimi peab olema unikaalne.

Tabel 2. Lapse pere mooduli valiku võimalused

Lapse pere lisainformatsioon	Valik
Lapse vanus	üksik
Pere suhtlemise keeled	mitmik, nimekirja valikuga
Kasvatuse koht	üksik
Pere tüüp	üksik
Lapse järjekorranumber peres	üksik
Pere elamise koht	üksik, nimekirja valikuga
Eestis elamise aeg (aastates)	üksik
Lapsevanema hariduse tase	üksik

Allikas: TÜ Narva Kolledži vene keele lektoraadi informatsiooni alusel lõputöö autori poolt koostatud.

Pärast isikliku teabe informatsiooni mooduli võib lisada info perekonna kohta. Selles moodulis on erinevad väljatüübid andmete sisestamiseks:

- väljad, mis peavad sisaldama ainult ühte piiratud valikuga väärtust;
- üks väärtus valikuga loendist oma variandi lisamise võimalusega;
- mitu väärtust valikuga loendist oma variandi lisamise võimalusega.

Perekonnainfo mooduli väljad ja väärtuse valimise meetodid on toodud **Tabel 2**.

Järgmine on moodul, kus sisestatakse info lapse õppetingimuste kohta. Kui laps käib praegu lasteaias, siis täidetakse lasteaiaga seotud teabeväljad. Kui laps õpib koolis, siis täidetakse kooliga seotud väljad. Nagu eelmises moodulis, teabe täitmiseks või lisamiseks on olemas erinevad viisid. Info valitavate väljade kohta on esitatud **Tabel 3**.

Tabel 3. Lapse õppimise tingimuste mooduli valiku võimalused

Lapse õppimise tingimused	Valik
Lasteaed	
Grupp	üksik
Gruppi tüüp	üksik, nimikirja valikuga
Lasteaia kaasabi	mitmik, nimikirja valikuga
Õppekavavälised tegevused	mitmik, nimikirja valikuga
Kool	
Klass	üksik
Klassi tüüp	üksik
Lasteaia kaimine	üksik
Õppimise programm	üksik, nimikirja valikuga
Kooli kaasabi	mitmik, nimikirja valikuga
Õppekavavälised tegevused	mitmik, nimikirja valikuga

Allikas: TÜ Narva Kolledži vene keele lektoraadi informatsiooni alusel lõputöö autori poolt koostatud.

Pärast lapse lisamist kuvatakse tema nimi teksti lisamise lehel. Lisatud laste arv pole piiratud. Samuti on võimalik laps nimekirjast eemaldada. Kustutamisenupu vajutamisel kuvatakse kinnitusaken.

Kasutuslugu 5.3. Teksti metaandmete lisamine

Peakasutajad: Toimetajad, administraator.

Eeltingimused: Kasutaja autoriseerimine süsteemis.

Kasutuslugu:

Teksti lisainfo sisestamise moodul sisaldab infot teksti salvestamise tingimuste kohta. Väljade väärtuste valimine toimub nagu eelmistes moodulites. On olemas erinevad viisid väljade täitmiseks väärtustega:

- väljad, mis peavad sisaldama ainult ühte piiratud valikuga väärtust;
- üks väärtus valikuga loendist oma variandi lisamise võimalusega;
- mitu väärtust valikuga loendist oma variandi lisamise võimalusega.

Teave väljade ja väärtuste lisamise võimaluste kohta on näidatud **Tabel 4**.

Tabel 4. Teksti lisainformatsiooni valiku võimalused

Teksti informatsioon	Valik
Üles kirjutamise tingimused	
Linn	üksik, nimikirja valikuga
Koht	üksik, nimikirja valikuga
Kuupäev	üksik
Aasta	üksik
Tekst	

Kõne vorm	üksik
Kõne erim	üksik
Kõne vaade	üksik
Kõne tüüp	mitmik, nimikirja valikuga
Stiil	üksik
Žanr	mitmik, nimikirja valikuga
Teksti teema	mitmik, nimikirja valikuga
Stiimul	üksik
Kõne ette valmistuse tase	üksik
Allkorpus	üksik

Allikas: TÜ Narva Kolledži vene keele lektoraadi informatsiooni alusel lõputöö autori poolt koostatud.

Kasutuslugu 5.4. Teksti sõnavormide märgistus

Peakasutajad: Toimetajad, administraator.

Eeltingimused: Kasutaja autoriseerimine süsteemis.

Kasutuslugu:

Pärast teksti salvestamist andmebaasi avaneb sõnavormi märgistuse aken. Selles aknas kuvatakse kogu varem lisatud tekst. Märgistusega sõnavormid kuvatakse mustana, ilma märgistuseeta hallina.

Märgistustöödega alustamiseks tuleb vajutada sõnavormile, mis on vaja märgistada. Pärast vajutamist ilmub moodul, kus kuvatakse valitud sõnavorm kogu märgistusega. Juhul, kui andmebaasis on sama märgistatud sõnavorm juba olemas, siis kuvatakse selle märgistus. Vajadusel saab sõna muuta või kustutada tekstist. Sõna muutmisel muudetakse ka märgistus, kui antud sõna on juba andmebaasis olemas. Kui sõna puudub andmebaasis,

siis tähistatakse sõna märgistamata olekuga. Märgistamiseks on vaja vajutada nuppu „Разметить“. Pärast vajutamist avaneb märgistusleht, kus hetkel süsteemi arenduse etapil on olemas kaks valikut sõna märgistamiseks: „существительное“ või „глагол“.

Pärast nupule „Сохранить“ vajutamist sõnavorm märgistusega salvestatakse andmebaasi ja leht suletakse automaatselt. Pärast salvestamist värvitakse sõnavorm mustaks. Sõnavormide märgistus on piiratud toimetaja rolliga. Need kasutajad saavad märgistada ainult enda poolt sisestatud tekste. Administraatorid saavad märgistada kõikide kasutajate tekste.

3.6 Otsing

Kasutuslugu 6. Otsingu vaade

Peakasutajad: Registreeritud kasutajad.

Eeltingimused: Autoriseerimine süsteemis.

Kasutuslugu:

Kõnekorpusse esmase etappi arendamise ajal otsinguleht tavakasutajate jaoks on peidetud. Otsingulehele sisenedes kuvatakse tekstide loend sorteeritud lisamise kuupäeva järgi kahanevalt, s.t. lähima lisamisajaga tekstid on üleval. Ühel lehel kuvatakse korraga viie teksti osaline vaade. Iga osa juures on viit teksti detailvaatele, mis sisaldab kogu informatsiooni, millest tuleb edasi juttu kasutusloos 6.2.

Otsingu tegemiseks lehel on tekstiväli, kus kasutaja saab sisestada sõna või lause. Otsingu alustamiseks tuleb vajutada nuppu „Найти“. Leht sisaldab ka filtrite moodulit, mille kasutamist kirjeldatakse allpool kasutusloos 6.1. Leitud tekstide arv kuvatakse otsimisnupu kõrval. Kui puuduvad täidetud otsinguväljad ja filtrid, kuvatakse andmebaasis tekstide olemasolev koguarv.

Kasutajaliidese eskiis:

LOGO RusLAPSED Добро пожаловать, Ольга! [Выход](#)

Главная **Поиск** Добавить текст Моя страница

Поиск текста по слову

[Использовать фильтр](#) [Найти](#)

10.11.2019
Lorem ipsum dolor sit amet et delectus accommodare his consul copiosae legendos at vix ad putent delectus delicata usu. Vedit dissentiet eos cu eum an brute copiosae hendrerit. Eos erant dolorum an. Per facer affert ut. Mei iisque mentitum moderatius cu...

07.11.2019
Lorem ipsum dolor sit amet et delectus accommodare his consul copiosae legendos at vix ad putent delectus delicata usu. Vedit dissentiet eos cu eum an brute copiosae hendrerit. Eos erant dolorum an. Per facer affert ut. Mei iisque mentitum moderatius cu...

01.11.2019
Lorem ipsum dolor sit amet et delectus accommodare his consul copiosae legendos at vix ad putent delectus delicata usu. Vedit dissentiet eos cu eum an brute copiosae hendrerit. Eos erant dolorum an. Per facer affert ut. Mei iisque mentitum moderatius cu...

25.10.2019
Lorem ipsum dolor sit amet et delectus accommodare his consul copiosae legendos at vix ad putent delectus delicata usu. Vedit dissentiet eos cu eum an brute copiosae hendrerit. Eos erant dolorum an. Per facer affert ut. Mei iisque mentitum moderatius cu...

25.10.2019
Lorem ipsum dolor sit amet et delectus accommodare his consul copiosae legendos at vix ad putent delectus delicata usu. Vedit dissentiet eos cu eum an brute copiosae hendrerit. Eos erant dolorum an. Per facer affert ut. Mei iisque mentitum moderatius cu...

< 1 2 3 4 5 6 7 114 115 116 117 118 119 120 >

Корпус детской речи RusLAPSED
© 2020

Andmed: Autori joonis.

Joonis 20. RusLAPSED teksti lisamise lehe vaade.

Kasutustlugu 6.1. Otsingu filtrite kasutamise

Peakasutajad: Registreeritud kasutajad.

Eeltingimused: Autoriseerimine süsteemis.

Kasutustlugu:

Otsingut võib teha mitte ainult sõna ja lause kaudu, vaid ka filtrite abiga. Saab kasutada ka ainult filtreid. Esimeses arengu etapis võetakse tööle nimekiri (lisa 3, tabel 3), mis sisaldab 5 väärtust. Pärast filtrite valimist tuleb vajutada nuppu „Найти“. Päringu tulemus kuvatakse otsingulehel tekstide loendina.

Kasutajaliidese eskiis:

The wireframe illustrates the RusLAPSED user interface. At the top, a header bar contains a 'LOGO' placeholder, the site name 'RusLAPSED', a user greeting 'Добро пожаловать, Ольга!', and a 'Выход' (Logout) button. Below the header is a navigation menu with four items: 'Главная' (Home), 'Поиск' (Search), 'Добавить текст' (Add text), and 'Моя страница' (My page). The 'Поиск' (Search) section features a text input field labeled 'Поиск текста по слову', a 'Использовать фильтр' (Use filter) button, and a 'Найти' (Find) button. A red message 'Строка с информацией о количестве найденных текстов' (Line with information about the number of found texts) is positioned above the filters. The 'Фильтрация по метаданным текста' (Filtering by text metadata) section includes two categories: 'Ребенок' (Child) with 'Пол' (Gender) and 'Возраст' (Age) dropdowns, and 'Текст' (Text) with 'Подкорпус' (Subcorpus), 'Разновидность речи' (Speech variety), and 'Тип речи' (Speech type) dropdowns. At the bottom, a date '10.11.2019' and a block of Lorem Ipsum placeholder text are shown.

Andmed: Autori joonis.

Joonis 21. RusLAPSED teksti otsingu vaade avatud filtritega.

Kasutuslugu 6.2. Teksti detailvaade

Peakasutajad: Registreeritud kasutajad.

Eeltingimused: Autoriseerimine süsteemis.

Kasutuslugu:

Detailvaate lehel saab näha täisteksti ja kogu lisateavet selle kohta. Kui vajutada laste kõne sõnavormile, siis saab näha sõna morfoloogilist märgistust, kui see on konkreetse sõna kohta tehtud. Lisainformatsioon teksti kohta kuvatakse lehe allosas ja see sisaldab ainult neid välju, mis selle teksti salvestamisel oli täidetud. Kui kasutaja on administraator või teksti lisanud toimetaja, saab ta selle teksti sõnad parandada või märgistada, vastavalt kasutusloole 5.4.

Kasutajaliidese eskiis:



Andmed: Autori joonis.

Joonis 22. *RusLAPSED täisteksti lehe vaade.*

4. Tehnoloogiate valik

Keelekorpuse RusLAPSED väljatöötamiseks kasutatavate tehnoloogiate valikul lähtuti lõputöö autori isiklikest kogemustest. Selles peatükis mainitud tehnoloogiatega sai autor kogemusi TÜ Narva Kolledži õppimise ajal, ettevõtte praktika ja töö ajal.

4.1 Angular 8

Angular – TypeScript'il põhinev platvorm üheleheliste veebirakenduste loomiseks. Peamised ehitusplokid Angular'is on NgModul'id, mis koguvad seotud koodi funktsioonide komplektidesse. Angular'i rakendus määratakse moodulite komplektiga ja see alati sisaldab vähemalt ühte peamoodulit, mille kaudu käib esmane laadimine. (Angular'i veebileht)

Koodi korraldamine erinevateks funktsionaalseteks mooduliteks võimaldab luua suuri ja keerukaid rakendusi ning kasutada samu mooduleid rakenduse erinevates kohtades. (Angular'i veebileht).

Rakendusmoodulites ühendatakse süsteemi erinevad komponendid ja teenused ning registreeritakse ka teekonnad komponentide käivitamiseks.

Süsteemi komponent määratleb klassi, mis sisaldab rakenduse loogikat ja andmeid. Komponendi andmemudel on seotud HTML lehega (tehakse kas ühepoolne või kahepoolne sidumine). Teenus sisaldab abifunktsioone (näiteks API poole pöördumine). Ühte teenust saavad kasutada ka mitu komponenti. (Angular'i veebileht)

4.2 ASP.NET Core Web API

ASP.NET Core – platvormist sõltumatu keskkond veebirakenduste loomiseks. Keskkond on lihtsasti integreeritav populaarsete kliendikeskkondade ja -teekidega, sealhulgas Angular ja Bootstrap teekidega. (Roth, Anderson jt: veebileht)

Korpuse RusLAPSED infosüsteemi serveriosa loomisel oli valitud lahenduseks ASP.NET Core Web API (*Application Programming Interface*). Web API on veebiteenused, millele pääsevad juurde teised rakendused (HTTP/HTTPS protokollide abil). See teenus töötab REST'i arhitektuuri baasil (esinduse oleku ülekandmine), see tähendab, et seda tüüpi päringuid kasutatakse järgmiselt: GET, POST, PUT, DELETE. Sarnast arhitektuuri kasutatakse sageli üheleheliste rakenduste ehitamiseks. (Metanit.com veebileht)

4.3 MS SQL Express ja Entity Framework Core.

MS SQL Express – Microsoft'i poolt arendatud andmebaasi haldussüsteem. Express versioon erineb MS SQL'i tavaversioonist selle poolest, et see on piiratud funktsionaalsusega ja tasuta, kuid RusLAPSED süsteemi jaoks on see täielikult sobilik.

Entity Framework – ORM (*object-relational mapping*) raamistik .NET keskkonna jaoks, mis on arendatud Microsofti poolt. Entity Framework võimaldab arendajale lihtsustada kõik tegevused objektidega, mis on seotud andmebaasipoolse tegevusega. Raamistik annab võimalust töödelda andmeid C# koodis kasutades objekte, mis on automaatselt seotud vastavate andmebaasi tabelitega. (Entity Framework Tutorial veebileht)

Infosüsteemi loomisel kasutati „Code First” lähenemist, mis tähendab, et kõigepealt loodi andmemudelid koodis ja nende põhjal Entity framework migratsioonide abil loodi automaatselt andmebaasi tabelid.

4.4 Bootstrap

Bootstrap – HTML, CSS ja JavaScript’i platvorm seadmetundliku disaini arendamiseks. Infosüsteemi huvitavamaks kujundamiseks kasutati Bootstrap’i teemat nimega „Spacelab“. Kõik selle teemalised materjalid on vabalt saadaval Bootswatch’i veebilehel, viit millele asub selle töö allikate loendis. Samuti töö kiiruse ja mugavuse jaoks oli kasutatud teek Bootstrap widgets, mis sisaldab valmis elemente Angular’i tööks (viit on allikate loendis).

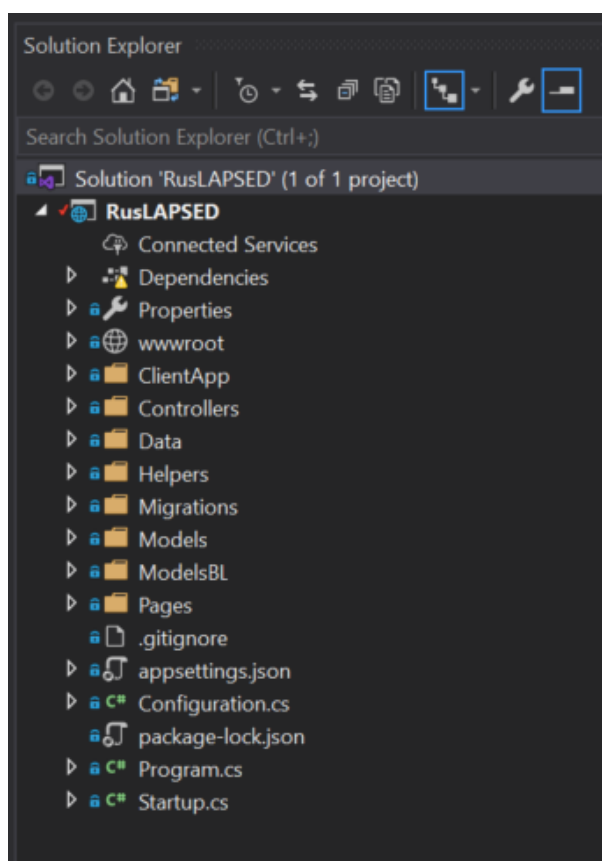
III. INFOSÜSTEEMI PROTOTÜÜBI LOOMINE

Infosüsteemi kasutajaliidese disaini ja funktsionaalsust muudeti ja täiendati töö käigus väga olulisel määral. Selles peatükis kirjeldatakse nende muudatuste ja täienduste tegemise käiku.

1. Rakenduse arhitektuurne lahendus

1.1. Projekti üldehitus

Projekti arhitektuurne lahendus on määratud valitud tehnoloogiatega. Projekti malli loomisel olid automaatselt genereeritud kõik peamised süsteemifailid, mille abil süsteem töötas. Töö alustamiseks korpuse infosüsteemi projektiga kustutati mõned automaatselt loodud failid (näidisveebilehed jms). Projekti praegust üldist ülesehitust võib näha **Joonis 23**.



Andmed: Autori ekraanipilt.

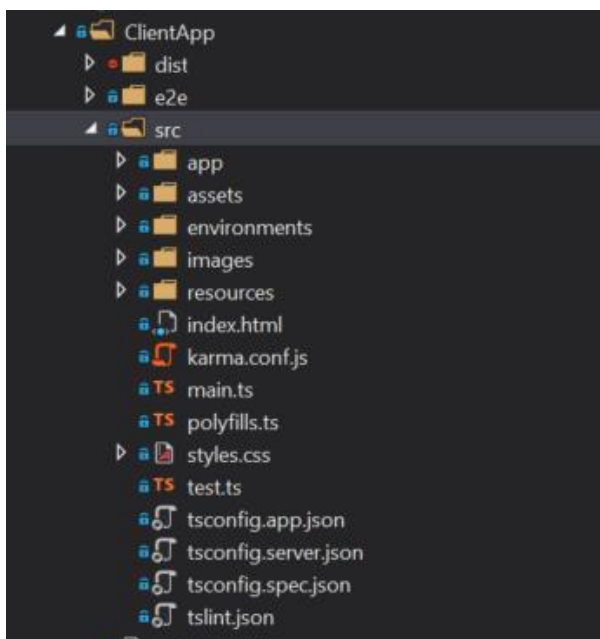
Joonis 23. Visual Studio. Projekti sisu üldvaade

Projekti autori muudetud või lisatud failid ja kaustad (kaustad on esile tõstetud paksu kirjaga):

- **ClientApp** – kaust sisaldab projekti kogu kliendi osa. Kausta sisu kirjeldatakse järgmises peatükis.
- **Controllers** – kaustas asuvad kontrollid, st klassid, mis sisaldavad meetodeid, mille abil võib teha HTTP/HTTPS päringuid. Neid kontrollereid kasutades toimub kliendi ja rakenduse serveri vaheline koostöö.
- **Data** – kaustas on kaks faili:
 - `ApplicationDbContext.cs` – Entity Framework'i klass, mis loob seost mudeli ja andmebaasi vahel.
 - `SeedData.cs` – sisaldab meetodit, mis programmi käivitamisel täidab automaatselt andmebaasi lähteandmetega, kui määratud andmed seal puuduvad.
- **Helpers** – kaust abiklassidega:
 - `Cipher.cs` – sisaldab paroolide krüptimise meetodeid.
 - `Enums.cs` – sisaldab kogu projektis kasutatud andmeid tüübiga Enum.
 - `Mail.cs` – sisaldab e-kirjade saatmise meetodeid.
 - `Resource.resx` – projekti serveripoolne ressursifail.
- **Migrations** – kaustas on kõik projektis tehtud migratsioonid.
- **Models** – projekti andmemudelid:
 - andmemudelid, mida kasutatakse andmebaasi töötamiseks
 - kaust **DTO** (*data transfer object*) – sisaldab andmemudeleid, mida kasutatakse suhtlemiseks rakenduse kliendi osaga.
- **ModelsBL** – sisaldab klasse ja liideseid ärilooikaga.
- `appsettings.json` – see fail sisaldab rakenduse konfigureeritavaid funktsioone, näiteks teekond andmebaasini.
- `Configuration.cs` – klass juurdepääsu lihtsustamiseks konfigureeritavatele andmetele, mis asuvad `appsettings.json` failis.
- `Startup.cs` – klass sisaldab protseduure, mida tuleb ära teha projekti käivitamiseks, näiteks andmebaasi konfigureerimine. Samuti selles failis registreeritakse klassid ja liideseid mudelite ärilooikaga.

1.2. Infosüsteemi kliendiosa

Infosüsteemi kliendiosa asub kaustas **ClientApp** > **src**, mille sisu võib näha *Joonis 24*.



Andmed: Autori ekraani pilt.

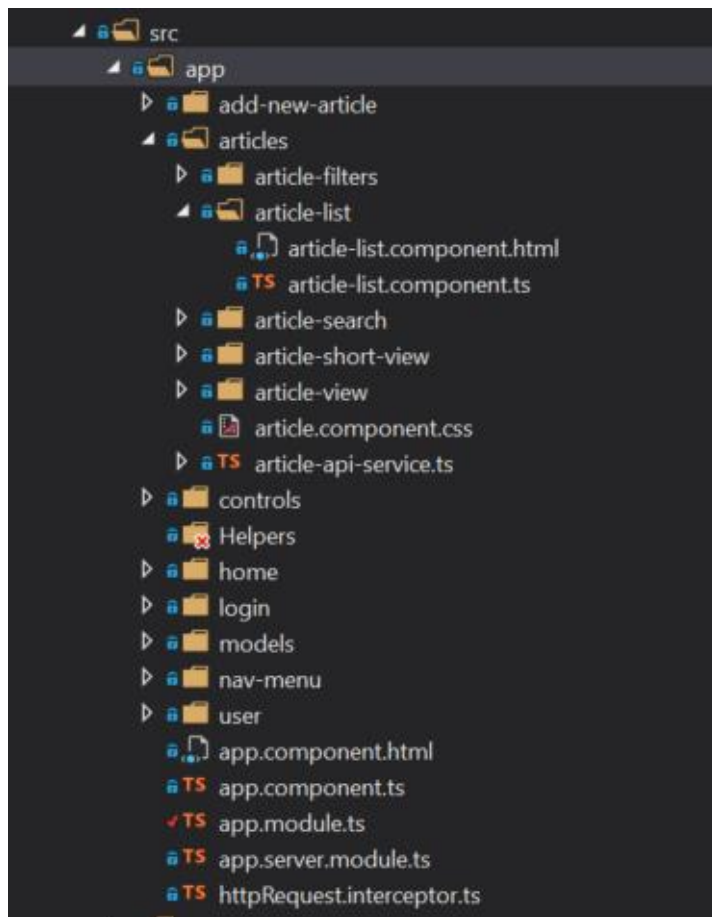
Joonis 24. Visual Studio. ClientApp kausta sisu üldvaade.

Kausta peamised failid, mida projekti autor muutis või lisas:

- **app** – kaust sisaldab peamoodulit ja süsteemi komponente. Kausta täpsemat sisu kirjeldatakse lõputöö järgmises alapeatükis.
- **images** – kaust sisaldab projekti päise pilti.
- **resources** – kaust sisaldab teenust resources.service.ts, mida kasutatakse kogu kasutajaliidese teksti kuvamiseks.
- index.html – fail on infosüsteemi leht, sisaldab viidet süsteemi põhikomponendile (<app-root>), mis asub kaustas nimega app.component.ts.
- styles.css – fail sisaldab infosüsteemi globaalseid kujunduslaade, nimelt Bootswatch Specelabi teemat ja mõnda autori polt lisatud kujundust.

1.2.1. Infosüsteemi projekti moodul ja komponendid

Kaust **app** sisaldab põhimoodulit ja kogu projekti komponente. Kausta sisu võib näha *Joonis 25*.



Andmed: Autori ekraani pilt.

Joonis 25. Visual Studio, **app** kausta sisu üldvaade.

Kausta **app** sisu:

- **add-new-article** – sisaldab faile, mis on rakendatud uue teksti lisamiseks andmebaasi.
- **articles** – kaust, mis sisaldab faile teksti kuvamiseks ja otsimiseks. Kasutades selles kaustas olevaid faile, kirjeldatakse komponentide loomise üksikasjalikumat arhitektuuri. Kõikide teiste moodulite komponentidel on sama struktuur nagu selle kausta komponentidel. Iga komponent sisaldab .html faili (lehe mall) ja .ts faili (klass andmetega ja komponentide funktsioonidega). Kaust sisaldab järgmisi alamkaustasid (**Joonis 25**):
 - **article-filters** – tekstiotsingu filtri komponent.
 - **article-list** – komponent, mis vastutab tekstide loendi kuvamise eest.
 - **article-search** – otsingu komponent.
 - **article-short-view** – komponent ühe tekstiosa kuvamiseks.
 - **article-view** – komponent, mis vastutab täisteksti kuvamise eest.

- `article.component.css` – fail, mis sisaldab selle mooduli laade.
- `article-api-service.ts` – teenus, millega tehakse päringuid programmi serveripoole.
- **controls** – kaust sisaldab kõiki komponente, mida kasutatakse rakenduse erinevates osades. Näiteks veateate kuvamise komponent (**error-messages-component**). Seda komponenti kasutatakse süsteemi kõikides osades, seetõttu komponent on kirjutatud eraldi failis, mis võimaldab seda rakenduse erinevates kohtades uuesti kasutada.
- **home** – avalehe komponente sisaldav kaust.
- **login** – sisselogimise mooduli sisaldav kaust.
- **models** – kaust sisaldab süsteemimudeleid, mida kasutatakse suhtlemiseks programmi serveriosaga. Selle kausta mudelid genereeritakse automaatselt, kasutades Visual Studio arenduskeskkonna lisandit nimega Typewriter. Mudelite automaatne loomine põhineb mudelitel, mis asuvad projekti serveriosa kaustas **DTO**.
- **nav-menu** – kaust peamenüü komponendiga.
- **user** – kaust registreerunud kasutajate isikliku lehe komponentidega.
- `app.component.html` – põhikomponendi mall, mis sisaldab viiteid projekti muudele komponentidele.
- `app.component.ts` – klass andmete ja komponentidega.
- `app.module.ts` – projekti põhimoodul, kus deklareeritakse kõik komponendid, teekonnad ja teenused.
- `httpRequest.interceptor.ts` – klass töötab kasutaja autoriseerimise teabega.

2. Andmemudel

Selles arenguetapis sisaldab andmemudel 27 erinevat olemit ehk tabelit, millest 9 on vahepealsed. Andmemudel sisaldab kahte olemit (*ArticleReviewers* ja *Messages*), mida ei ole antud töös kasutatud ja mis on sinna lisatud projekti järgmisi arendusetappe silmas pidades. Täielikku andmemudelit seostega võib näha lisa 4.

2.1. Olemite metaandmed

Paljud olemid sisaldavad samu välju metaandmetega. Kuna iga olemit loomiseks on vaja eraldi C# klassi, siis selleks, et mitte dubleerida neid välju kõikides nendes klassides, löi

autor klassi nimega *MetaData*, mis oli emaklassiks kõikidele nendele klassidele, nii et kõik nad pärisid klassist *MetaData* selles sisalduvaid välju. Klass sisaldab järgmisi välju:

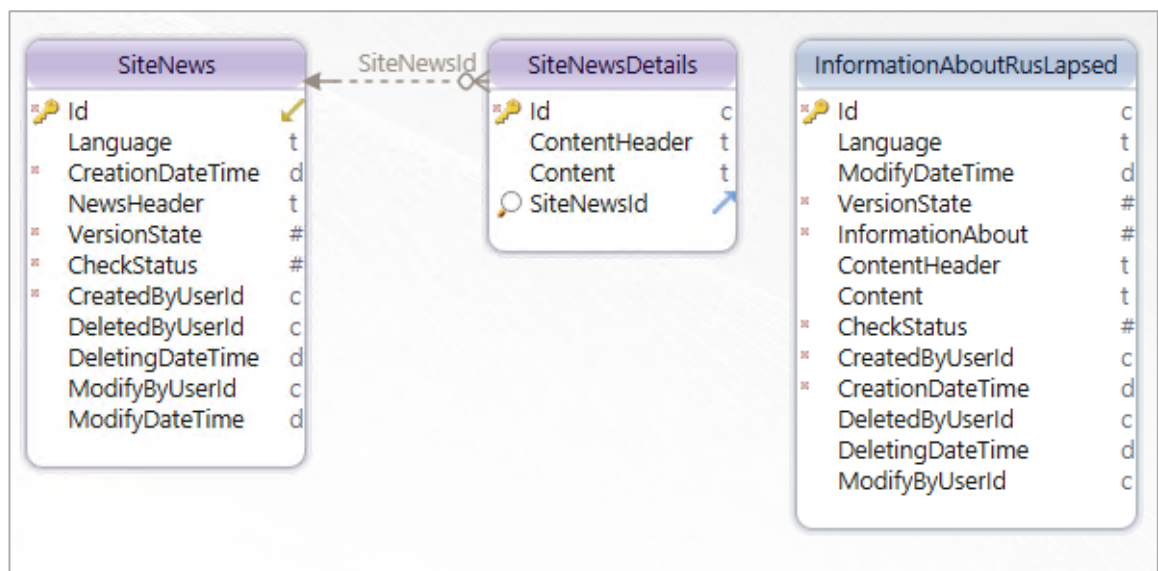
- *CheckStatus* on Enum'i tüüpi andmeväli, mis sisaldab teavet andmete kontrollimise kohta teiste kasutajate poolt. See väli oli loodud objekti oleku jälgimiseks ja ta võib sisaldada järgmist teavet: ootel, tagastatud, kinnitatud, kustutatud. Andmebaasis on see väli täisarv.
- *CreationDateTime* – objekti loomise kuupäev.
- *CreatedByUserId* – looja identifitseerimistunnus.
- *ModifyDateTime* – muutmise kuupäev.
- *ModifyByUserId* – objekti muutnud kasutaja identifitseerimistunnus.
- *DeletingDateTime* – kustutamise kuupäev.
- *DeletedByUserId* – objekti kustutanud kasutaja identifitseerimistunnus.

Süsteem on üles ehitatud nii, et andmebaasist ei kustutata peaaegu ühtegi objekti, vaid peidetakse kasutaja eest. Objekti peitmiseks registreerimata kasutaja eest oli tehtud väljad *DeletingDateTime* ja *DeletedByUserId*. Samuti mõne objekti kohta on vaja salvestada versioone, selleks loodi väljad *ModifyDateTime* ja *ModifyByUserId*.

2.2. Uudiste ja korpuse info tabelid

Kogu korpuse teave asub tabelis *InformationAboutRusLapsed* (**Joonis 26**). See olem sisaldab teavet korpuse ja arendajate kohta, kuvatakse vastavalt süsteemi moodulites „О корпусе“ ja „О разработчиках корпуса“. Tabel sisaldab järgmisi välju:

- *Id* – identifitseerimistunnus.
- Informatsioon objektist (väljad, mis pärinevad *MetaData* klassist).
- *Language* – tekst, mis sisaldab salvestatud teabe keelt.
- *VersionState* – väli, mis sisaldab objekti olekut. Andmebaasis on see täisarv, C# klassis on see andmetüüp Enum. See väli võib sisaldada järgmist objekti olekut (versiooni): praegune, muudetud, kustutatud.
- *InformationAbout* – selles väljas märgitakse, kas teave selles kirjes käib projekti või arendajate kohta.
- *ContentHeader* – tiitli tekst.
- *Content* – tekst.



Andmed: Autori ekraanipilt.

Joonis 26. Andmemudeli osa tabelitega avalehe infoga.

Tabel *SiteNews* (**Joonis 26**) sisaldab korpuse uudiseid, koosneb järgmistest väljadest:

- *Id* – identifitseerimistunnus.
- Informatsioon objektist (väljad pärinevad *MetaData* klassist).
- *Language* – tekst, mis sisaldab salvestatud teabe keelt.
- *VersionState* – väli, mis sisaldab objekti olekut (versiooni).
- *NewsHeader* – uudiste pealkiri.

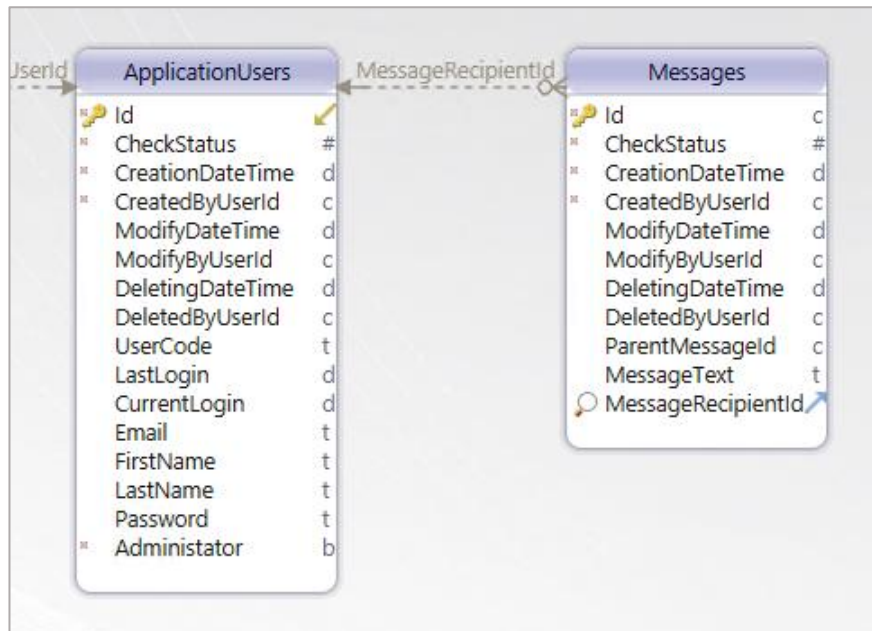
Üks postitus korpuse uudiste lehel võib sisaldada uudiste plokki piiramata uudiste arvuga, mis kogutakse tabelis *SiteNewsDetails* (**Joonis 26**).

2.3. Kasutajate ja kasutaja sõnumite tabelid

Teavet korpuse kõigi kasutajate kohta hoitakse tabelis *ApplicationUsers*, mis on kuvatud **Joonis 27**. Samuti joonisel on kuvatud sõnumite tabel *Messages*. Sõnumite olem luuakse privaatsõnumite saatmiseks teistele kasutajatele. Esimeses arenduse etapis seda funktsionaalsust ei lisata. Sõnumite tabel sisaldab järgmisi välju:

- *Id* – identifitseerimistunnus.
- Informatsioon objektist (väljad pärinevad *MetaData* klassist).

- *ParentMessageId* – väli eelmise sõnumi identifikaatori jaoks. Selle välja abil saab rekursiivselt laadida andmebaasist kõik sõnumite ahelad.
- *MessageText* – sõnumi tekst.
- *MessageRecipientId* – sõnumi adressaadi identifitseerimise väli. Vajalik kasutaja viite määramiseks.



Andmed: Autori ekraanipilt.

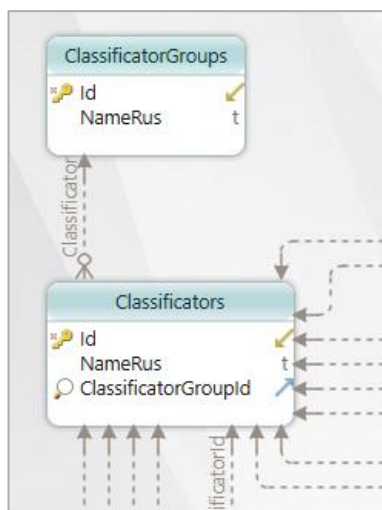
Joonis 27. Andmemudeli osa. Kasutajate ja sõnumite tabelid.

Tabel *ApplicationUsers* (**Joonis 27**) koosneb järgmistest väljadest:

- *Id* – identifitseerimistunnus.
- Informatsioon objektist (väljad pärinevad *MetaData* klassist).
- *UserCode* – tekst kasutaja järjekorranumbriga.
- *LastLogin*, *CurrentLogin* – väljad sisaldavad viimase ja praeguse sisselogimise kuupäevi.
- *Email* – tekst, mis sisaldab e-posti aadressi, mida kasutatakse kasutajatunnuseks.
- *FirstName*, *LastName*, *Password* – tekst eesnime, perekonnanime ja parooliga.
- *Administrator* – Boolean andmetüüpi väli kasutajarolli määratlemiseks.

2.4. Klassifikaatorite ja žanrite tabelid

Mõne mudeli väljad võivad sisaldada loendeid, mida süsteemi kasutaja vajadusel uuendab ja täiendab. Nendest loenditest pärinevatel andmetel on sama struktuur, mistõttu otsustati kasutada klassifikaatorite (*Joonis 28*) ja žanrite (*Joonis 29*) olemeid.



Andmed: Autori ekraanipilt.

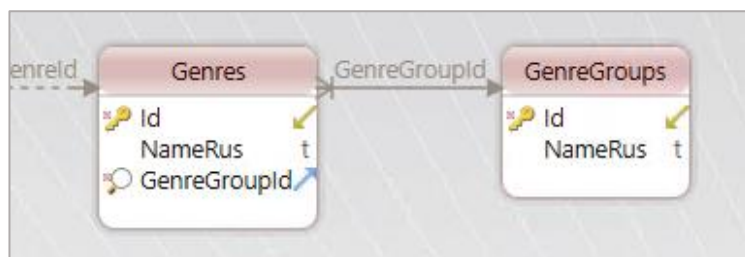
Joonis 28. Andmemudeli osa klassifikaatorite tabelitega.

Tabel *ClassifierGroups* koosneb kahest väljast (*Joonis 28*):

- *Id* – identifitseerimistunnus, mis andmebaasis on täisarv ja C# klassis on Enum'i andmetüüp. See väli sisaldab teavet selle kohta, millist tüüpi rühma klassifikaator kuulub.
- *NameRus* – tekst, mis sisaldab klassifikaatorirühma nime vene keeles.

Tabel *Classifiers* sisaldab järgmisi välju (*Joonis 28*):

- *Id* – identifitseerimistunnus.
- *NameRus* – tekst, mis sisaldab klassifikaatori nime vene keeles.
- *ClassifierGroupId* – määrab, kas klassifikaator kuulub konkreetseesse rühma, mis vastab mudeli *ClassifierGroups* väärtusele.



Andmed: Autori ekraanipilt.

Joonis 29. Andmemudeli osa žanrite tabelitega.

Klassifikaatorite ja žanrite (**Joonis 29**) olemi andmetel on sarnane struktuur. Erinevus on selles, et žanri grupe on väga palju, mitte üks, nagu klassifikaatori grupis. Andmebaasist teabe võtmise mugavuse huvides otsustati kasutada erinevaid tabeleid.

Kasutajaliidese eesti keelde tõlkimisel tuleb neisse olemisse lisada väljad eestikeelse nime salvestamiseks.

2.5. Peamine andmemudeli osa teksti tabeliga

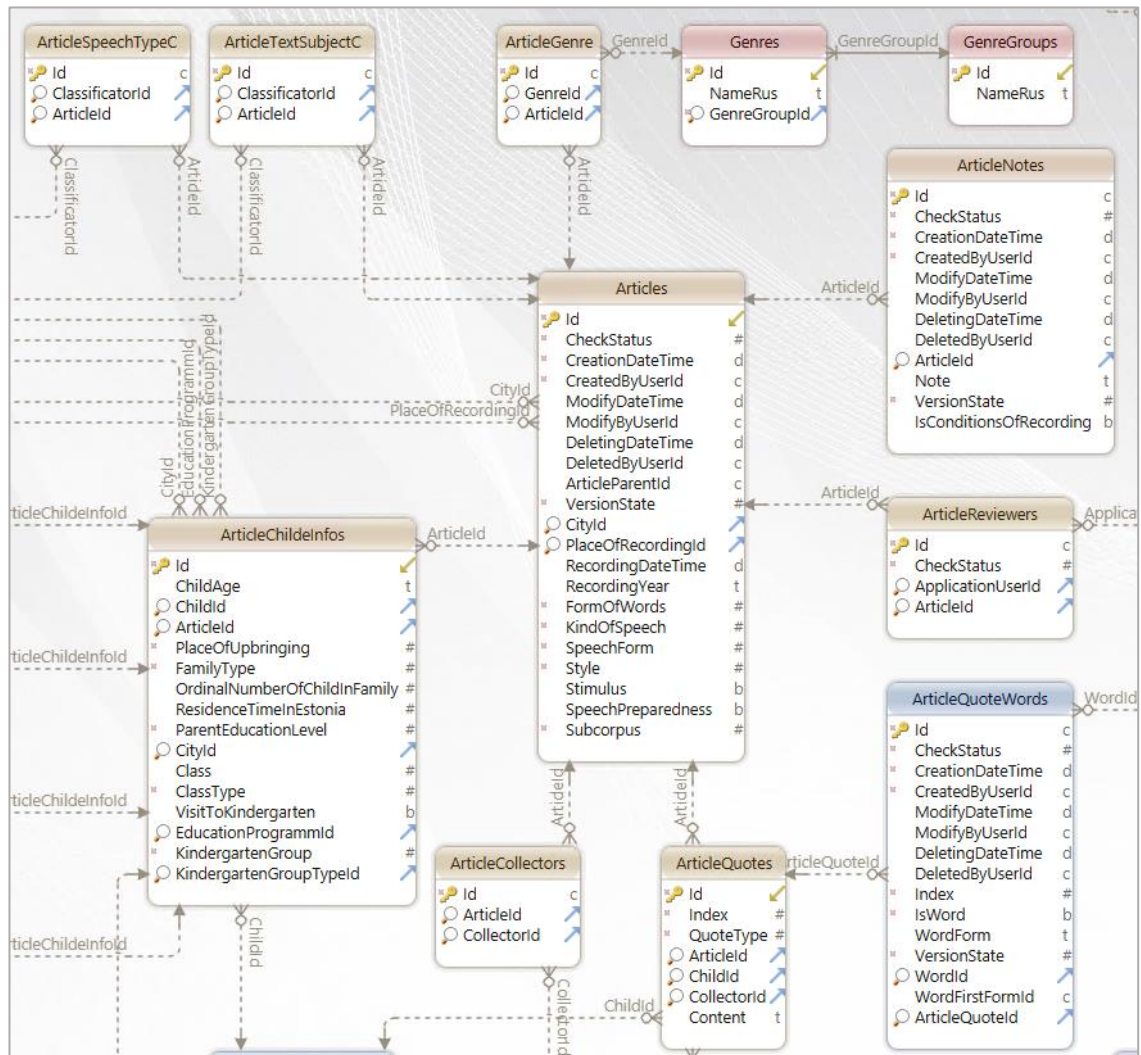
Süsteemi põhitabel on teksti olemi tabel *Articles* (**Joonis 30**), mis sisaldab teavet teksti- ja salvestustingimuste kohta ning millel on palju seoseid andmemudeli teiste olemitega.

Tabel *Articles* sisaldab:

- *Id* – identifitseerimistunnus.
- Informatsioon objektist (väljad pärinevad *MetaData* klassist).
- *ArticleParentId* – teksti eelmise versiooni identifitseerimistunnus. See seadistatakse programmiselt, kui tekst pärast salvestamist oli muutunud. Vajalik tekstide versiooniajaloo loomiseks.
- *VersionState* – väli, mis sisaldab objekti olekut (versiooni).
- *CityId*, *PlaceOfRecordingId* – klassifikaatori identifitseerimistunnus, andmed teksti salvestamise tingimustest.
- *RecordingDateTime* – teksti salvestamise kuupäev.
- *RecordingYear* – teksti salvestamise aasta, on vajalik konkreetse kuupäeva kohta teabe puudumise juhul.
- *FormOfWords*, *KindOfSpeech*, *SpeechForm*, *Style*, *Subcorpus* – teksti metaandmed, C# klassis väli andmetüübiga Enum.

- *Stimulus, SpeechPreparedness* – teksti metaandmed, väli andmetüübiga Boolean.

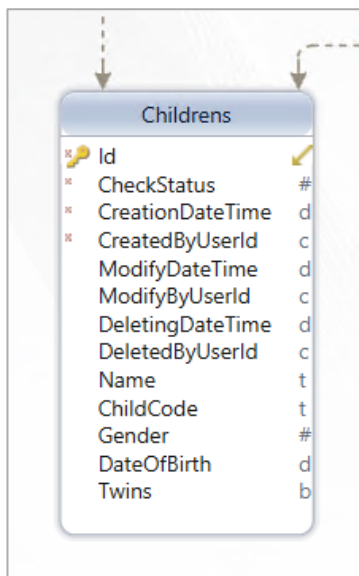
Joonis 30 võib näha vahetabeleid (*ArticleSpeechTypeC*, *ArticleTextSubjectC*, *ArticleGenre*) teksti ning eelpool nimetatud klassifikaatorite ja žanrite tabelite seostamiseks mitu-mitmele seose tüübiga.



Andmed: Autori ekraanipilt.

Joonis 30. Andmemudeli osa tabeliga *Articles* ja selle seostega mudeli teiste tabelitega.

Iga tekst võib sisaldada mitme lapse kõnet, kellel on omakorda erinevad isikuandmed. Nende andmete jaoks loodi tabel *ArticleChildeInfos*. Üks tekst võib viidata mitmele olemile. Osa lapse isikuandmeid ei saa muuta, seetõttu otsustati muudetud andmed salvestada tabelisse *Childrens*, mille väljad on näidatud **Joonis 31**. Tabel *ArticleChildeInfos* on seotud tabeliga *Childrens*.

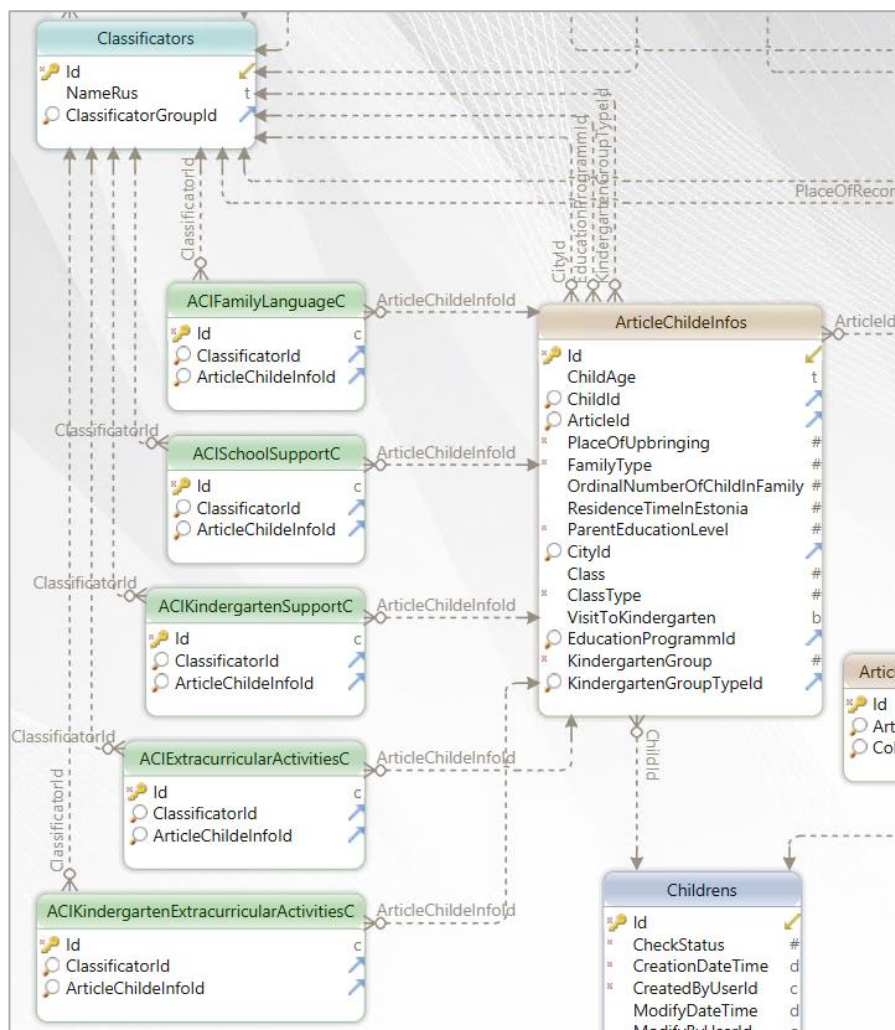


Andmed: Autori ekraanipilt.

Joonis 31. Andmemudeli osa laste tabeliga.

Tabel *ArticleChildeInfos* on seotud tabeliga *Classificators* kahel erineval viisil: kolm välja on seotud seosega üks-mitmele ja viis välja – seosega mitu-mitmele läbi vahetabelite, mis on märgitud **Joonis 32** rohelise värviga. Tabel sisaldab järgmisi välju:

- *Id* – identifitseerimistunnus.
- *ChildId* – lapse identifitseerimistunnus, kelle kohta info sisestatakse.
- *PlaceOfUpbringing*, *FamilyType*, *ParentEducationLevel*, *ClassType*, *KindergartenGroup* – lisaandmed lapse kohta, C# klassis on see väli andmetüübiga Enum.
- *OrdinalNumberOfChildInFamily*, *ResidenceTimeInEstonia*, *Class* – lisaandmed lapse kohta, andmetüüp on täisarv.
- *CityId*, *EducationProgramId*, *KindergartenGroupId* – klassifikaatorite identifitseerimistunnused.
- *VisitToKindergarten* – teksti metaandmed, väli andmetüübiga Boolean.



Andmed: Autori ekraanipilt.

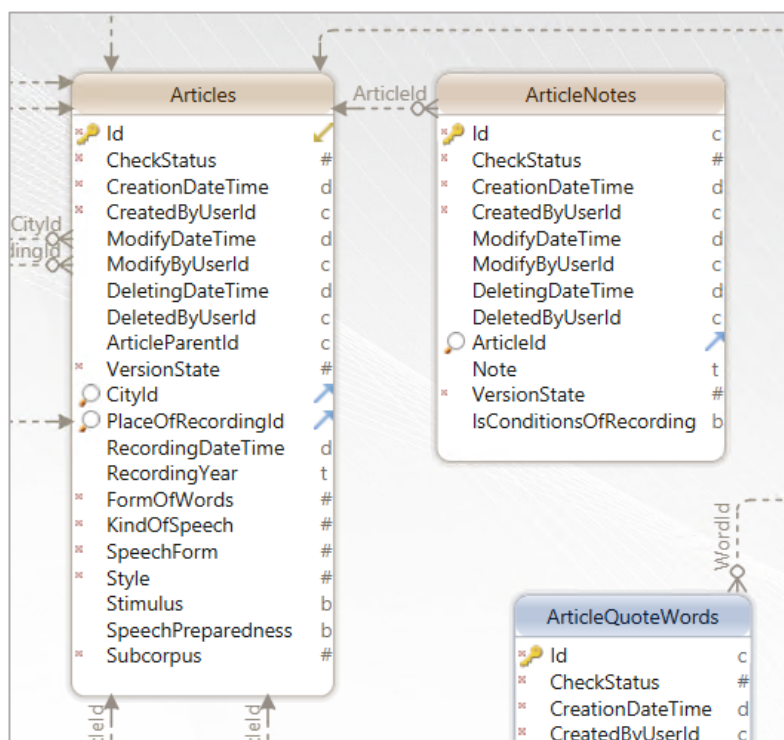
Joonis 32. Andmemudeli osa tabeliga *ArticleChildeInfos* koos seostega teiste tabelitega.

Iga tekst võib sisaldada palju kommentaare. Üheks kommentaari liigiks on kommentaar teksti salvestamise või lapsele antud ülesande tingimustest. Seda kommentaari kuvatakse alati koos tekstiga. Muud kommentaarid lisatakse teksti toimetaja poolt (kasutaja, kes kontrollib teksti sisu) ja need ei tohi olla registreerimata kasutajale kättesaadavad. Esimeses arenduse etapis tekstide kontrollimise ja kommenteerimise funktsionaalsust ei lisatud.

ArticleNotes kommentaaride salvestamise tabel sisaldab järgmisi välju (**Joonis 33**):

- *Id* – identifitseerimistunnus.
- Informatsioon objektist (väljad pärinevad *MetaData* klassist).
- *ArticleId* – teksti identifitseerimistunnus.
- *Note* – kommentaari tekst.

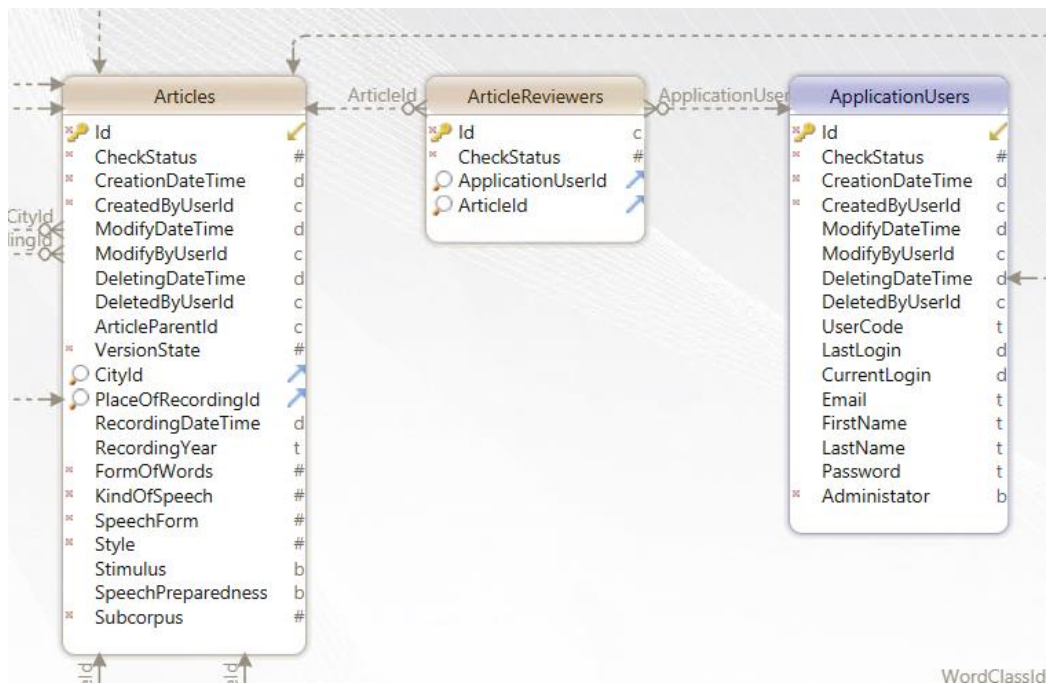
- *VersionState* – kommentaari versiooni seis.
- *IsConditionsOfRecording* – väli andmetüübiga Boolean, milles märgitakse, kas on tegemist toimetaja kommentaariga või teksti salvestamise tingimustega.



Andmed: Autori ekraanipilt.

Joonis 33 Andmemudeli osa tekstide ja kommentaaride tabelitega.

Teiste kasutajate teksti kontrollimiseks tuleb tekstile määrata toimetajad. Esimeses arenduse seda funktsionaalsust ei lisatud. Toimetajate määramisel seostatakse teksti olem toimetaja olemiga tabeli *ArticleReviewers* abil, mis on sisuliselt vahetabel tekstide ja kasutajate tabelite sidumiseks omavahel seosega mitu-mitmele. Tabelisse *ArticleReviewers* on lisatud väli *CheckStatus* teksti kontrollimise oleku jälgimiseks. Nagu eelnevalt oli mainitud, C# klassis on *CheckStatus* Enum'i andmetüüp. Tabel *ArticleReviewers* on toodud **Joonis 34**.



Andmed: Autori ekraanipilt.

Joonis 34 Andmemudeli osa tekstide ja kasutajate tabelite seostega.

Tekst salvestatakse andmebaasi osade kaupa. Iga tekst on *ArticleQuotes* olem (**Joonis 35**). See olem sisaldab teavet tekstimooduli tüübi kohta (*QuoteType*), nimelt teavet selle kohta, kas tekstimoodul on lapse või koguja kõne. Tabelis on järgmised väljad:

- *Id* – identifitseerimistunnus.
- *Index* – tekstimoodulite kuvamise järjekord.
- *QuoteType* – teksti tüüp.
- *ChildId*, *CollectorId* – lapse ja koguja identifitseerimistunnused.
- *Content* – tekst.

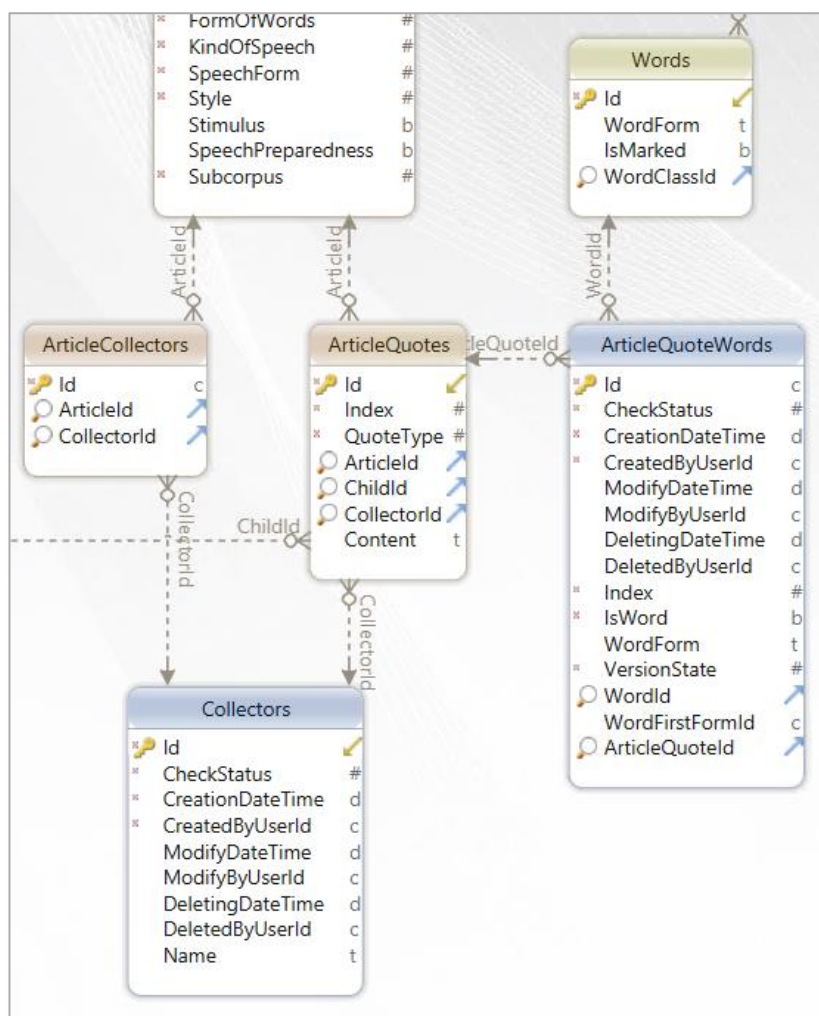
Olem *ArticleQuotes* on seotud tabeliga *ArticleQuoteWords*, mis sisaldab ainult laste sõnu. Tabel *ArticleQuoteWords* on mõnes mõttes vahepealne tabel, mis seob seosega mitmitmele tabelleid *Words* ja *ArticleQuotes*, kuid see samuti sisaldab ka lisateavet. Tabel sisaldab mitte ainult sõnavorme, vaid ka mitmesuguseid tekstis sisalduvaid märkusi ja kirjavahemärke.

Tabel *ArticleQuoteWords* sisaldab järgmisi välju:

- *Id* – Identifitseerimistunnus.
- Informatsioon objektist (väljad pärinevad *MetaData* klassist).

- *Index* – sisu kuvamise järjekord.
- *IsWord* – väli andmetüübiga Boolean, mis näitab kas tegemist on sõnavormiga või mitte.
- *WordForm* – tekst sõnavormi või märkidega.
- *VersionState* – objekti olek.
- *WordId* – sõna identifitseerimistunnus.
- *WordFirstFormId* – sõna identifitseerimistunnus, viide sõnale, mis on selle sõna algvormiks.

Sõna algvormile viite lisamise funktsionaalsust rakenduse esimeses arendusetapis ei oli tehtud.



Andmed: Autori ekraanipilt.

Joonis 35. Andmemudeli osa tabeli *ArticleQuotes* seostega teiste tabelitega.

Tabel *Words* (**Joonis 35**) sisaldab järgmisi välju:

- *Id* – identifitseerimistunnus.
- Informatsioon objektist (väljad pärinevad *MetaData* klassist).
- *WordForm* – sõnavorm.
- *IsMarked* – väli andmetüübiga *Boolean*, näitab kas sõnal on märgistus või mitte.
- *WordClassId* – klassifikaatori identifitseerimistunnus.

3. Kliendi ja serveri osade suhtlemine

Rakenduse kliendi ja serveri pool töötavad ühtemoodi viisil, mis on kirjeldatud allpool. Kliendi ja serveri poolte vahelist kommunikatsiooni kirjeldatakse kogujate nimekirja mooduli näitel.

Selles moodulis kuvatakse andmebaasis olevate kogujate loend. Kui avada kausta **ClientApp>src>app>add-new-article>article-collectors**, võib näha kahte faili, mis vastutavad koguja lisamise mooduli kuvamise eest kasutajaliideses. Fail `article-collectors.component.html` sisaldab kuvatava lehe malli. Fail `article-collectors.component.ts` sisaldab andmeid ja funktsioone. Samas failis on ühendus vajalike teenustega. Näiteks, programmi HTML-kood andmebaasis sisalduva kogujate loendi kuvamiseks on **Joonis 36**, real 50 on viide teenusele nime kuvamiseks kasutajaliideses valitud keeles. See teenus ühendatakse `article-collectors.component.ts` faili konstruktoris, failiühendust näidetakse **Joonis 37**, koodi rea number 34.

HTML-i faili 52-54 ridadel (**Joonis 36**) kuvatakse kogujate loend, mis deklareeritakse ts-failis (**Joonis 37**) real 23.

```
49
50 <label>{{resourcesService.getTranslate('Collector-name')}}</label>
51 <select class="form-control" (change)="setCollectorFromBase($event)">
52   <option *ngFor="let collector of collectorsFromBase" value="{{collector.id}}">
53     {{collector.name}}
54   </option>
55 </select>
56
```

Andmed: Autori ekraanipilt.

Joonis 36 Visual Studio. Näidis kliendiosa HTML-faili koodist.

Joonis 37 (rea 38) võib näha meetodit `ngOnInit()`. See meetod käivitatakse automaatselt mooduli laadimisel. Praeguses näites see meetod kutsub meetodit `getCollectors()`, mis vastutab juurdepääsu eest andmebaasist andmete saamise teenusele meetodi `getAllCollectorsFromBase()` abil. Viimase meetodi sisendiks on otsingutekst. Kui otsingut ei olnud tehtud, siis see parameeter on tühi ja see tähendab, et väärtus ei mõjuta tulemuste kitsendamist.

```
32
33 constructor(
34     public resourcesService: ResourcesService,
35     private addArticleApiService: AddArticleApiService,
36     private modalService: NgbModal) { }
37
38 ngOnInit() {
39     this.getCollectors();
40 }
41
42 getCollectors() {
43     this.addArticleApiService.getAllCollectorsFromBase(this.searchString).then(result => {
44         this.collectorsFromBase = result;
45     })
46 }
47
```

Andmed: Autori ekraanipilt.

Joonis 37. Visual Studio. Näide kliendiosa `.ts` faili koodist.

Teenusklass sisaldab kõiki rakenduse serveripoolse juurdepääsu meetodeid, mis võimaldab projekti koodi korras hoida ja eristada API pöördumiste loogikat komponendi loogikast. Serveriosa poole suunatakse GET- ja POST-päringud, teave edastatakse andmetüübi „string“ või JSON viisil (DTO-andmemudelite abil). Meetod serverile juurdepääsuks on kuvatud **Joonis 38**, koodi real 25. See meetod viitab kontrolleri, mille nimi algab sõnaga „Collector“.

```
23
24 }
25 getAllCollectorsFromBase(partOfName: string) {
26     return this.http.get<CollectorDTO[]>(this.baseUrl + 'collector?partOfName=' + partOfName)
27         .toPromise()
28         .then(response => <CollectorDTO[]>response);
29 }
30
```

Andmed: Autori ekraanipilt.

Joonis 38. Visual Studio. Kliendiosa `AddArticleApiService` klassi koodi osa.

Rakenduse serveri poolel kõik kontrollid on paigutatud kausta **Controllers**. Üldtõttud näide viitab kontrolleri `CollectorController.cs`, kus on meetod `GetAllCollectors()`, mis omakorda viitab meetodile, mis asub *Collector* mudeli

ärioloogika klassis (kaust **ModelsBL**, klass **CollectorBL.cs**). Andmebaasi poole otse kontrolleriiga pöördumist peetakse „halvaks tavaks“. Kui kontrolleriis asub suur kogus koodi, see muutub raskesti mõistetavaks ja arusaadavaks, seetõttu on kombeks eraldada kontrollereid ja rakenduse ärioloogikat. Sellega seoses serveri poolel oli loodud kaust süsteemi ärioloogikaga, mis sisaldab liideseid ja nende klasse funktsioonidega.

4. RusLAPSED veebilehe disain

Korpuse disaini määrab valitud teek – Bootstrap – ja selle lisaarendused: teema nimega Spacelab ja lisandmoodulite kogu Bootstrap widgets. Lisaks kasutati korpuse kujundamisel FontAwesome'i veebilehelt võetud ikoonide tasuta versioone ja korpuse päise pilt oli võetud tasuta piltide kogust Pexels. Veebilehte võib näha viite kaudu, mis on toodud selle peatüki 5. osas.

Enamik veebilehe funktsioone on praegu kättesaadavad vaid registreeritud kasutajale, seetõttu saab registreerimata kasutaja näha vaid avalehte. Kõike veebilehe elemente võib näha piltidel, mis on toodud lisas 5.

4.1. Veebilehe disaini muutused

Projekti käigus on veebilehe disain muutunud ja praegu erineb algsetest eskiisidest. Mõned muudatused olid tingitud kasutatavatest teekidest.

Üks suurimaid muudatusi on toimunud teksti lisamise lehel. Algselt eeldati, et süsteemi lisatakse kohe kogu tekst, kuid arendusprotsessi käigus sai selgeks, et see pole mugav lahendus, kuna lisamisel erinevaid teksti osi on vaja erinevalt töödelda (kogujatekst ja laste tekst). Sellega seoses otsustati lisada tekst moodulite kaupa, eraldi väljaga sõnade autori märkimiseks (*Joonis 39*).

Текст

Комментарии, информация о ситуации записи текста

< Колю попросили рассказать, чем он занимался в гостях. >

Содержание статьи

[Коля, 7,4:]

Мы/ когда в прятки играли/ я знаешь где прятался?/Ну [...]ну в шкафу/ но только не на нижнем/ где большие дверцы/ а где маленькие дверцы наверху. Артурка меня поднял. Было не страшно/ но высоко. Ааа [...] мы играли/ звали ниточки/мы пили яблочный сок/мы играли в телефон и мы/еще/и мы играли в тайной комнате/и мы потом/и мы потом ели мороженое/ а до мороженого мы ели торт./И потом мы с Артуром танцевали./Дискотека была./ А потом/ мы/мы потом/после музыки играли в обезьянок-акробатов.

[Маша, :]

А дальше?

Блок текста

Автор слов

☒ Ребенок

☐ Собиратель

Цитата

Добавить

Andmed: Autori ekraanipilt.

Joonis 39. Korpuse RusLAPSED teksti lisamise vaade.

Samuti on pisut muutunud korpuse avaleht. Alguses oli plaanitud korpuse statistika moodul teha eraldi lehel, kuid hiljem otsustati ümber, kuna esimeses arendamisetapis see sisaldas liiga vähe teavet. Sellega seoses statistika moodul oli lisatud avalehele (**Joonis 40**). Tulevikus, kui üksikasjalikuma statistika loomiseks kogutakse piisavalt andmeid, siis võib minna tagasi algse idee juurde ja teha eraldi leht statistika jaoks.

69

Новости

О корпусе

О разработчиках корпуса

Новости

Lorem ipsum dolor sit amet

19.05.2020

Ut enim ad minim veniam

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Статистика корпуса

количество текстов: 1

количество

словоупотреблений: 62

Andmed: Autori ekraanipilt.

Joonis 40. Korpuse RusLAPSED avalehe vaade.

Mõned muudatused on toimunud samuti kasutaja isiklikul lehel. Sellel lehel otsustati loobuda külghenüüst ja paigutada viited isikliku teabe muutmiseks kompaktsemal kujul (**Joonis 41**).

Добро пожаловать, Olga Gandshu!

Персональный код в инфосистеме:

Дата и время последнего входа: 19.05.2020, 11:58

Редактировать личные данные

Изменить пароль

Личный кабинет

Поиск

Таблица пользователей

Код	Имя	Фамилия	Дата и время последнего входа	Статус
Olga	Gandshu	19.05.2020, 13:12	Действующий	<div>Детальный просмотр</div>

Добавить нового пользователя

Andmed: Autori ekraanipilt.

Joonis 41. Korpuse RusLAPSED isiklikulehe vaade.

5. Viited lõputöö olulistele lisadele

Infosüsteemi RusLAPSED kood: https://dev.azure.com/olgagandshu/_git/RusLAPSED-open%20project?path=%2FRusLAPSED

Korpuse RusLAPSED veebilehelt (asub TÜ sisevõrgus): <http://172.26.0.46/>

IV. JÄRGMISED SAMMUD INFOSÜSTEEMI ARENDAMISEKS

1. Töö sõnavormiga

Järgmine suur osa arengust on sõnavormide üksikasjaliku märgistamise võimaluse loomine. See märgistus peaks olema üles ehitatud samal põhimõttel nagu teksti metaandmete märgistamine, see tähendab märgistamine pakutud võimaluste valimisega.

Praegu ei saa salvestada andmebaasi kahte identset sõnavormi. See on ajutine lahendus, mis oli valitud selle lihtsuse tõttu, kuna mitme ühesuguse sõnavormiga lahenduse väljatöötamine ei olnud selle töö ülesannete hulgas. Tulevikus aga sellise võimaluse peab kindlasti lisama, kuna vene keeles samadel sõnavormidel võivad olla erinevad tähendused ja märgistused. Võimalikuks lahenduseks on luua olemasoleva sõnavormi dubleerimise võimalus selle edasiseks märgistamiseks. Täpsete märgistusnõuete saamiseks on vaja pöörduda TÜ Narva Kolledži vene keele lektoraadi poole.

2. Töö teksti kontrollimisega

Infosüsteemi arendamise esimesel etapil ei ole loodud funktsionaalsust, mis võimaldaks teistel kasutajatel kontrollida ühe kasutaja poolt lisatud teksti. Teksti lisanud kasutaja saab teha vigu, mistõttu on vaja luua võimalus saata tekst kontrollimiseks teistele kasutajatele. Selle jaoks on andmebaasis juba loodud tabel *ArticleReviewers* (vt andmemudeli kirjeldust osas III, peatükis 2.5).

Lõputöö autori mõtte kohaselt teksti parandaja võib kirjutada kommentaare teksti kohta ja tagastada tekst koos kommentaaridega teksti andmebaasi lisanud kasutajale või võib kinnitada, et tekstis vigu ei ole. Peale teksti korrektsuse kinnitamist teksti saab näidata otsingulehel olevate tekstide loendis, kontrollimata teksti ei tohi otsingulehel näidata. Praegu loendis näidatakse kõik tekstid, sest kontrollimise funktsionaalsust pole veel loodud.

3. Teksti otsing sõna järgi

Esimeses arenduse etapis tekstide otsimine metaandmete järgi on rakendatud osaliselt. Rakenduse kliendi osas on tehtud tekstide filtreerimise võimalus vastavalt 5 väärtusele, ehkki andmebaas sisaldab palju rohkem teksti metaandmeid. Algselt oli kavas luua

võimalus filtreerida kõigi metaandmete järgi, kuid esimeses arenduse etapis tuli sellest loobuda. Otsingu mooduli eskiis koos kõikide filtritega oli tehtud, kuid antud töös seda ei kasutatud (lisa 6).

4. Kasutajaliidese mitmekeelsus

Korpuse loomise peamine eesmärk on võimalus uurida vene keelt rääkivate laste kõnet. Sellega seoses on kasutajaliidese peamiseks keeleks tehtud vene keel. Tulevikus eestikeelse kasutajaliidese lisamise mugavuse huvides autor lisas projekti vajalikud ressursifailid ja kasutajaliidese tõlkimise eest vastutava meetodi. Ressursikaustade asukoha võib leida käesoleva lõputöö 3. peatükist.

Samuti uudistel ja korpuse kirjeldusel andmebaasis on tehtud väli keele jaoks. See väli on loodud selleks, et kuvada uudiseid teises keeles. Uudiste täitmisega tegelevad veebilehe administraatorid, seetõttu on vaja luua võimalus lisada uudiseid ka eestikeelsena. Selle saab teha, näiteks, kasutades praegust liidese keelt või luues kasutajale võimaluse uudiste keel käsitsi märkida.

5. Muu funktsionaalsus

Selle lõputöö uurimise osas tehtud võrdleva analüüsi käigus oli välja toodud teiste korpuste huvitavad võimalused, mida saab rakendada ka antud tööd projekti järgmistes arenduse etappides:

- sagedusgraafikute koostamine;
- teksti alla laadimine kasutaja seadmesse;
- võimalus vaadata kõiki sõnavorme märgistusega, mis on olemas andmebaasis;
- juhised veebilehega töötamiseks.

KOKKUVÕTE

Lõputöö eesmärk oli alustada laste kõne korpuse RusLAPSED infosüsteemi arendust TÜ Narva kolledži vene keele lektoraadi jaoks. See töö on korpuse arendamise esimene etapp ning autori põhiülesanne oli luua süsteemi alus ja lisada sellele peamine funktsionaalsus, mis võimaldaks süsteemi kasutajatel alustada tööd tekstide lisamisega süsteemi ja tekstide märgistamisega.

Enne süsteemi arenduse alustamist toimusid kohtumised vene keele lektoraadi esindajate ja üliõpilastega, kes uurivad laste kõnet ja koguvad materjale korpuse loomiseks. Kohtumiste käigus selgitati välja süsteemi peamised ootused. Võttes arvesse need ootused autor viis läbi teiste keelekorpuste uurimise ja tegi võrdleva analüüsi. Uuring näitas nende korpuste tugevaid külgi ning need teadmised olid kasutatud nõuete sõnastamiseks ja infosüsteemi RusLAPSED arhitektuuri loomiseks.

Pärast nõuete välja selgitamist ja kinnitamist tellija poolt, autor pani kirja esmased kasutajalood ja lõi veebilehe kasutajaliidese eskiisid. Hiljem, kohtumiste käigus tellijaga, kasutajalugudesse ja süsteemi disaini tehti vajalikud muudatused ja täiendused.

Tehnoloogiate valik antud projekti jaoks tugines autori teadmistele ja varasematele kogemustele, mis olid saadud ettevõttepraktika ja töö käigus. Infosüsteemi loomiseks otsustati kasutada järgmisi tööriistu: Angular 8, ASP.NET Core Web API, MS SQL Express, Entity Framework Core ja Bootstrap.

Kokkuvõttes, on korpuse RusLAPSED esimese arenguetapi lõpus loodud: avaleht; leht teksti otsimiseks ja kuvamiseks; võimalus lisada ja märgistada uut teksti ja sõnade metaandmeid; võimalus muuta isikuandmeid. Töö kirjalikus osas on põhjalikult kirjeldatud projekti struktuur ja toodud andmebaasi üksikasjalik kirjeldus. Töö viimases peatükis loetletakse järgmised võimalikud arenduse etapid. Need kirjeldused võivad aidata teisi arendajaid jätkata tööd laste kõne RusLAPSED korpuse projekti arendamisega.

Lõputöö autor arvab, et kõik antud töös püstitatud ülesanded said edukalt täidetud. Järgmiste kuude jooksul plaanib autor jätkata infosüsteemi toetamist ja vigade parandamist.

SUMMARY

The goal of this thesis was to begin development of the information system RusLAPSED for children's speech corpus, which is a work in progress for the division of Russian language and literature of Narva College of Tartu University. This work is the first stage in the development of the corpus and the main task for the author was to create the foundation for the project and add the main functionality, which would allow the users of the system to begin work on filling the system with texts and marking them up.

Before the beginning of the development, there were meetings with representatives of the division of Russian language and literature and their students, who study the speech of children and collect materials for the corpus. During the meetings, the author collected and wrote down the main expectations for the system, after which she could proceed with conducting research and comparative analysis of other language corpora. The study revealed the strengths and useful features of other corpora and the author used this knowledge to formulate requirements and development strategy for the RusLAPSED information system.

When the scientists from division of Russian language and literature confirmed the requirements for the system, the author collected user stories and created site design sketches. Later, after more meetings with the scientists, who initiated this project, the author adjusted the user stories and the design of the system.

The author made her choice of technologies for this project based on her personal knowledge and experiences, which she got from the enterprise practice and later work. The author decided to create the information system using such tools as Angular 8, ASP.NET Core Web API, MS SQL Express, Entity Framework Core, and Bootstrap.

At the end of this stage of the project development, the corpus has the main page, a page for searching and displaying texts, the ability to add new text and markup their metadata, the ability to markup words, and the ability to edit personal data. In the written part of the work, the author added a detailed description of the database and the structure of the project. The final chapter of the work provides a list of further possible stages of development of the project. These descriptions can help future developers to proceed with next stages of the development of the children's speech corpus RusLAPSED.

The author is confident that she successfully solved and implemented all tasks, which were set up for this work. Over the next months, the author will continue to support and debug the information system.

KIRJANDUS

Arvuti maailm. Muischnek, K.; Orav H. (2002) *Eesti keele korpused ja arvutileksikonid — mis on olemas ja mida veel vaja on*. https://www.am.ee/3859_5F0.HTM (viimati vaadatud 20.02.2020)

Muischnek, K.; Orav H.; Kaalep H.; Õim H. (2003) *Eesti keele tehnoloogilised ressursid ja vahendid*. Haridus- ja Teadusministeerium, Eesti keelenõukogu, Eesti Keele Sihtasutus, Tallinn
<https://www.cl.ut.ee/yllitised/keeletehnoloogilised+ressursid.pdf> (viimati vaadatud 23.02.2020)

Brown korpuse kasutus juhend = Brown Corpus Manual. Kucera, H.; Francis, W.N. (1964, 1979) *Manual of information to accompany a Standard Corpus of Present-Day Edited American English, for use with Digital Computers*. Providence, Rhode Island, Department of Linguistics, Brown University
<http://korpus.uib.no/icame/brown/bcm.html#tc> (viimati vaadatud 24.02.2020)

Brown korpus = The Brown Corpus (1998) W3-Corpora project
https://www1.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html (viimati vaadatud 24.02.2020)

LOB korpuse kasutus juhend = *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with Digital Computers* (1978)
Department of English University of Oslo
<http://korpus.uib.no/icame/manuals/LOB/INDEX.HTM> (viimati vaadatud 24.02.2020)

Eesti keele koondkorpus (2018) <https://www.cl.ut.ee/korpused/segakorpus/> (viimati vaadatud 24.02.2020)

etTenTen <http://www.keeleeveeb.ee/dict/corpus/ettenten/about.html> (viimati vaadatud 24.02.2020)

Saksa keele instituudi korpus = IDS, Das Leibniz-Institut für Deutsche Sprache (2020) *Korpora geschriebener Gegenwartssprache*. <https://www1.ids-mannheim.de/kl/projekte/korpora/> (viimati vaadatud 10.03.2020)

EVKK = Tallinna Ülikooli eesti vahekeele korpus <http://evkk.tlu.ee/> (viimati vaadatud 10.03.2020)

CoRST. Corpus of Russian Student Texts (2015) http://web-corpora.net/learner_corpus/ (viimati vaadatud 18.03.2020)

VKRRK = Vene keele rahvuslik korpus. *Национальный корпус русского языка* (2020) <http://www.ruscorpora.ru/new/corpora-team.html#prog> (viimati vaadatud 20.03.2020)

Git. About <https://git-scm.com/about> (viimati vaadatud 08.05.2020)

SQL = Microsoft. *What is SQL Server Management Studio (SSMS)?* <https://docs.microsoft.com/en-us/sql/ssms/sql-server-management-studio-ssms?view=sql-server-ver15> (viimati vaadatud 08.05.2020)

Angular. *Introduction to Angular concepts* (2020) <https://angular.io/guide/architecture> (viimati vaadatud 13.05.2020)

Roth, D.; Anderson, R.; Luttin, S. (2020) Microsoft Docs. *Introduction to ASP.NET Core* <https://docs.microsoft.com/en-gb/aspnet/core/introduction-to-aspnet-core?view=aspnetcore-3.1> (viimati vaadatud 13.05.2020)

Metanit.com. *WEB API* (2019) <https://metanit.com/sharp/aspnet5/23.1.php> (viimati vaadatud 13.05.2020)

Entity Framework Tutorial. *What is Entity Framework?* <https://www.entityframeworktutorial.net/what-is-entityframework.aspx> (viimati vaadatud 13.05.2020)

Park, T. Bootswatch. *Spacelab* <https://bootswatch.com/spacelab/> (viimati vaadatud 13.05.2020)

Bootstrap widgets <https://ng-bootstrap.github.io/#/home> (viimati vaadatud 18.05.2020)

Font Awesome <https://fontawesome.com/> (viimati vaadatud 19.05.2020)

Pexels <https://www.pexels.com/photo/white-blank-notebook-733857/> (viimati vaadatud 19.05.2020)

LISAD

LISA 1. CoRST ja VKRK korpuste kasutajaliideste näited

Subcorpus

Period

from 2014

to 2015

Sex of author

☐ m ☐ f ☒ any

Course

- ☐ Check All
- ☐ 5 курс спец
- ☐ 1 курс спец
- ☐ 3 курс бак
- ☐ 4 курс бак
- ☐ 1 курс бак
- ☐ 1 курс бак
- ☐ 6 курс спец
- ☐ 2 курс спец
- ☐ 1 курс бак
- ☐ 2 курс бак
- ☐ 4 курс спец

Text features

Type of assignment	Academic major	Domain
<input type="checkbox"/> Check All	<input type="checkbox"/> Check All	<input type="checkbox"/> Check All
<input type="checkbox"/> Обзор конференции	<input type="checkbox"/> журналист	<input type="checkbox"/> политология
<input type="checkbox"/> курсовая	<input type="checkbox"/> филолог	<input type="checkbox"/> психология
<input type="checkbox"/> автобиография	<input type="checkbox"/> математик	<input type="checkbox"/> разнородная тематика
<input type="checkbox"/> вкр	<input type="checkbox"/> психолог	<input type="checkbox"/> менеджмент
<input type="checkbox"/> заявление	<input type="checkbox"/> лингвист	<input type="checkbox"/> журналистика
<input type="checkbox"/> эссе	<input type="checkbox"/> дизайнер	<input type="checkbox"/> право
<input type="checkbox"/> аннотация	<input type="checkbox"/> менеджер	<input type="checkbox"/> социология
<input type="checkbox"/> абзац	<input type="checkbox"/> логист	<input type="checkbox"/> экономика, социология
<input type="checkbox"/> семестровая работа	<input type="checkbox"/> историк	<input type="checkbox"/> Риторика
<input type="checkbox"/> аннотация проекта	<input type="checkbox"/> политолог	<input type="checkbox"/> социолингвистика
<input type="checkbox"/> диплом	<input type="checkbox"/> Мирэк	<input type="checkbox"/> риторика
<input type="checkbox"/> коммерческое предложение	<input type="checkbox"/> экономист	<input type="checkbox"/> лингвистика
<input type="checkbox"/> Эссе	<input type="checkbox"/> культуролог	<input type="checkbox"/> психология, педагогика
<input type="checkbox"/> реферат	<input type="checkbox"/> мирэк	<input type="checkbox"/> образование
<input type="checkbox"/> обзор конференции	<input type="checkbox"/> юрист	<input type="checkbox"/> филология
	<input type="checkbox"/> социолог	

Annotation

☐ Only checked ☐ Only with finished annotation

OK

Andmed: CoRST veebilehe autori ekraanipilt.

Joonis 1. CoRST korpus. Filtrid teksti otsimiseks.

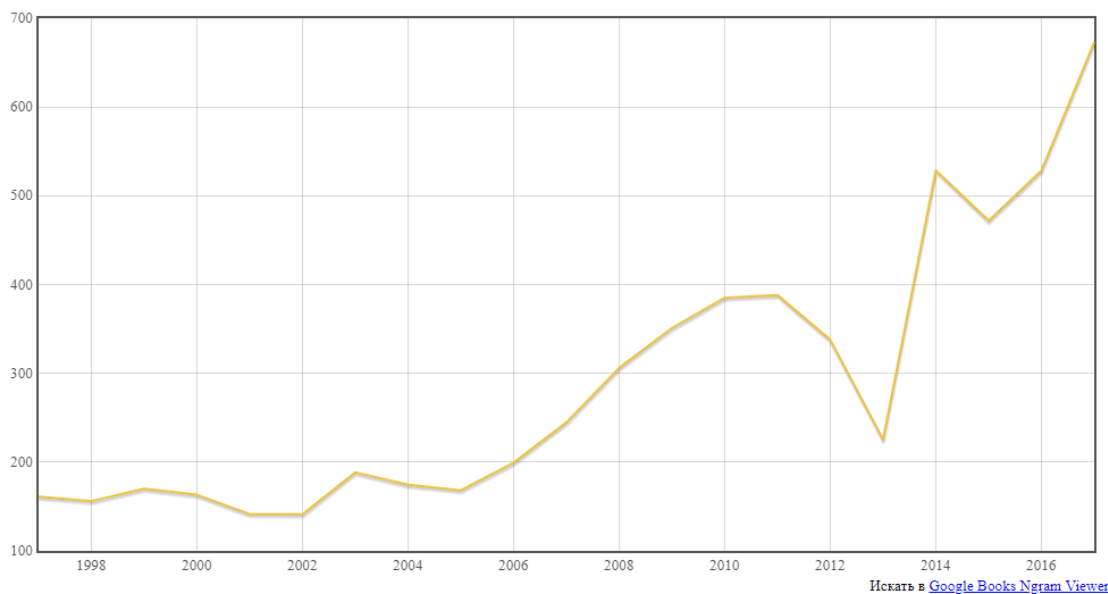


Распределение по годам в основном корпусе с 1997 по 2017 (сглаживание 2)

Слова или сочетания слов, через запятую:

Годы с по

Сглаживание:



[Показать таблицы](#) ▼

Искать в [Google Books Ngram Viewer](#)

Поиск осуществлен системой Яндекс. Поиск
При цитировании примеров просим ссылаться на Национальный корпус русского языка

Andmed: VKRK veebilehe autori ekraanipilt.

Joonis 2. VKRK korpus. Sõna „мама“ kasutamise sageduse graafik 1997-2017 aastate tekstides.

Выбор подкорпуса

Вы можете задать подмножество корпуса, по которому в дальнейшем будет вестись поиск. Подробнее о параметрах текста см. в разделе «[Параметры текста](#)».

Омонимия

☐ Только тексты со снятой грамматической омонимией [?]
☐ Только тексты с неснятой грамматической омонимией

Основные параметры текста [?]

Название
 Автор текста
 Пол: ☒ любой ☐ мужской ☐ женский
 Год рождения: с по ☐ Точное вхождение
 Год создания: с по ☐ Точное вхождение

Жанр и тип текста [?]

1. Художественные тексты ☐

Жанр текста [выбрать](#)

 Тип текста [выбрать](#)

 Место и время описываемых событий [выбрать](#)

2. Нехудожественные тексты ☐

Сфера функционирования [выбрать](#)

 Тип текста [выбрать](#)

 Место и время описываемых событий [выбрать](#)

2. Нехудожественные тексты ☐

Сфера функционирования [выбрать](#)

 Тип текста [выбрать](#)

 Тематика текста [выбрать](#)

Орфография

☐ Только тексты в старой орфографии
☐ Только тексты в новой орфографии

После выбора соответствующих параметров нажмите кнопку «Далее» и перейдите к просмотру списка документов, входящих в подкорпус.

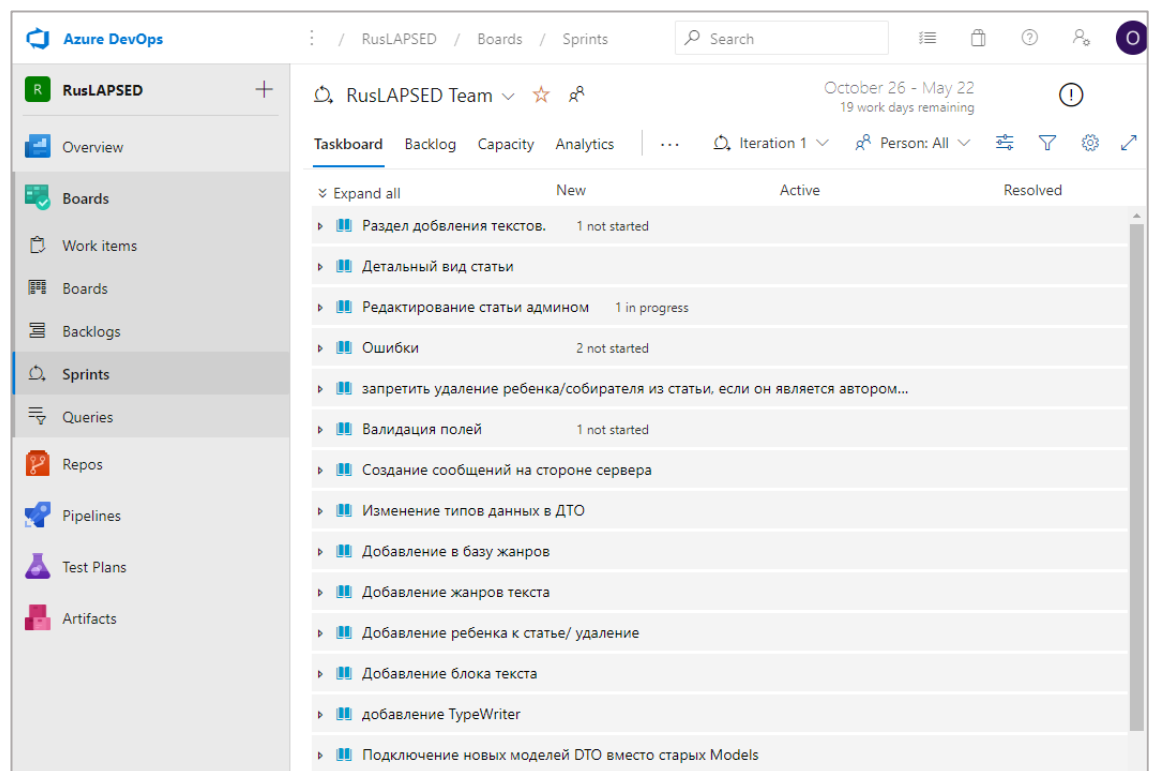
Национальный корпус русского языка
 © 2003–2020

Поиск осуществляется системой Яндекс.Поиск

Andmed: VKRK veebilehe autori ekraanipilt.

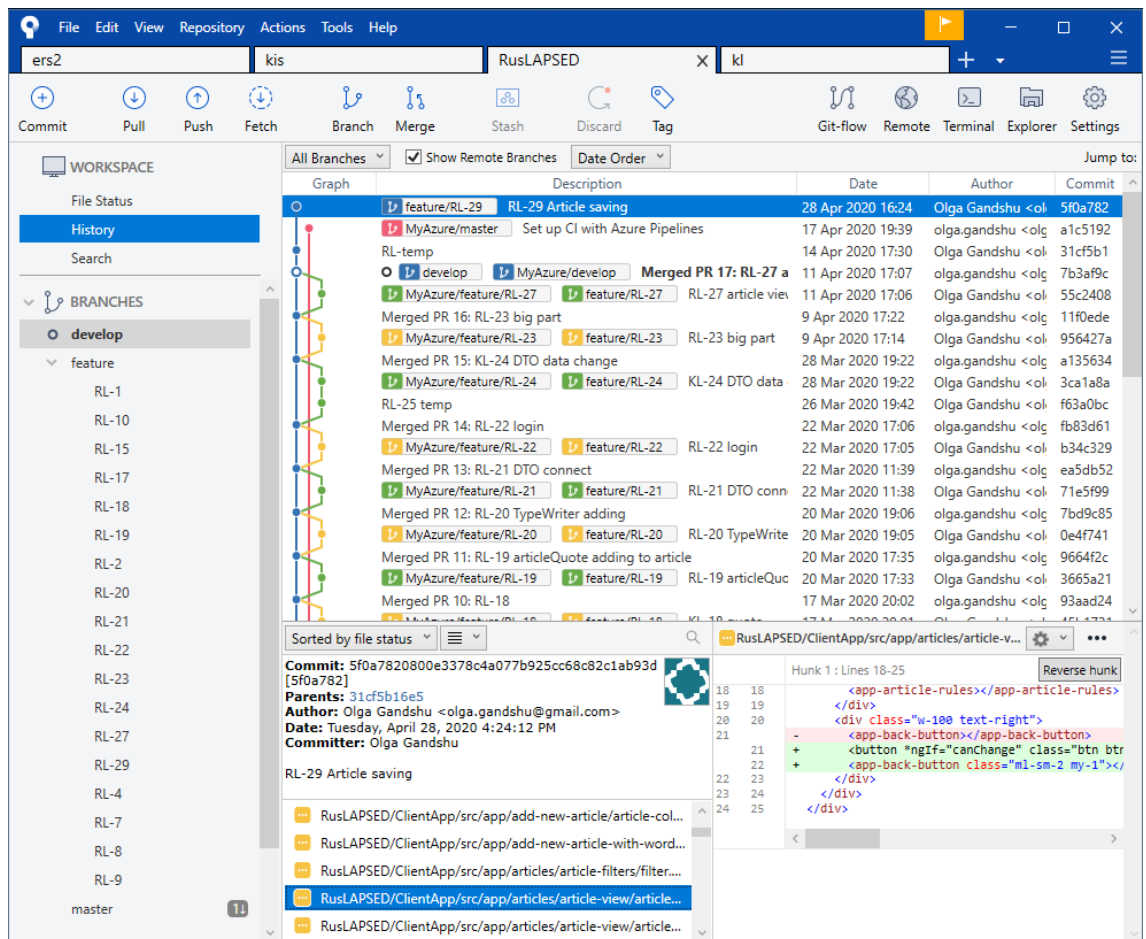
Joonis 3. VKRK korpus. Otsingu filtrid.

LISA 2. Projekti haldamise vahendite vaated



Allikas: Lõputöö autori ekraanipilt.

Joonis 1. Azure DevOps töölaud ülesannetega.



Allikas: Lõputöö autori ekraanipilt.

Joonis 2. Sourcetree töölaud.

LISA 3. Teksti metaandmete tabelid

Tabel 1. Lapse ja teksti metaandmete tabel

Ребенок	
Имя (условное)	
Код ребенка	
Пол ребенка	мужской/женский
Дата рождения	
Возраст на момент записи	дд.мм.гггг
Семья	
Язык(и) семейного общения	русский, эстонский, английский, украинский, немецкий, белорусский, шведский, финский.
Место воспитания	семья/ Дом ребенка
Тип семьи	полная/ неполная
Номер ребёнка в семье	первый/второй/третий и т. д.
Близнецы	да/ нет
Место жительства семьи	Нарва, Силламяэ, Йыхви, Кохтла-Ярве, Таллинн, Тарту, Нарва-Йыэсуу
Продолжительность жизни в Эстонии	
Уровень образования родителей	основное/среднее/среднее специальное/высшее
Условия обучения	
Школа	
Класс	порядковый номер
Тип класса	обычный/языковое погружение/специальный
Посещение ДДУ	да/нет
Программа обучения	государственная/ облегчённая (LÕK) / облегчённая (toimetulekuõpe) / облегчённая (hooldusõppekava)
Поддержка, оказываемая в учебном заведении	логопед/ спец. педагог/ социальный работник/ психолог
Внешкольная учебная деятельность	занятия с репетитором по русскому языку/ занятия с репетитором по другим предметам/ посещение логопеда
Детское дошкольное учреждение	
Группа	старшая младшая/ средняя/ старшая/ подготовительная
Тип группы детского дошкольного учреждения	языковое погружение(эстонский/английский)/ обычная группа/ логопедическая/ двустороннее языковое погружение
Поддержка, оказываемая в учебном заведении	логопед/ психолог
Внешкольная учебная деятельность	посещение кружков раннего развития/ посещение логопеда
Условия записи	

Место записи	школа г. Нарвы/ детское дошкольное учреждение г. Нарвы/ школа г. Силламяэ / детское дошкольное учреждение г. Силламяэ / школа г. Йыхви / детское дошкольное учреждение г. Йыхви / школа г. Кохтла-Ярве / детское дошкольное учреждение г. Кохтла-Ярве / школа г. Таллинна/ детское дошкольное учреждение г. Таллинна/ школа г. Тарту / детское дошкольное учреждение г. Тарту / школа г. Нарва-Йыэсуу/ детское дошкольное учреждение г. Нарва-Йыэсуу
Год записи	гггг
Время записи	дд.мм.гггг
Собиратель	
Текст	
Форма речи	устная/письменная
Разновидность речи	монолог/диалог/полилог
Вид речи	продуктивная речь/репродуктивная речь/продуктивно-репродуктивная речь
Тип речи	нарратив/рассуждение/описание
Стиль	нейтральный/сниженный
Жанр	<i>перечень жанров см. в отдельном документе; предусмотреть возможность пополнять перечень</i>
Тематика текста	<i>(возможность выбирать из выпадающего меню и пополнять его новыми темами)</i>
Стимул	да/нет
Степень подготовленности речи	подготовленная/неподготовленная
Объём (количество словоформ)	
Объём (письменная речь) (количество словоформ; количество предложений)	
Подкорпус	устный/письменный

Allikas: Tabel on koostatud TÜ Narva Kolledži vene keele lektoraadi esindajate ja tudengite poolt.

Tabel 2. Žanrite tabel.

Текст	
Жанр	Разговор: <ul style="list-style-type: none"> • бытовой • информационный разговор • разговор по душам • этикетный • болтовня • сплетня • разговор в игре (игра «рядом») • игра «в компании» • инсценированный диалог в сюжетно ролевой игре • разговор по телефону
	Жанры поля побуждения: <ul style="list-style-type: none"> • предложение (для разметчиков: сюда включается и приглашение) • совет (для разметчиков: сюда включаются и рекомендации) • требование • приказ совет (для разметчиков: сюда включаются и команды) • предупреждение • запрет • разрешение • упрашивание • уговоры • убеждения • жалобы
	Этикетные жанры: <ul style="list-style-type: none"> • благодарность • извинения • приветствие • прощание • поздравления
	Оценочные: <ul style="list-style-type: none"> • похвала • комплимент • одобрение • похвальба • обвинение • осуждение • порицание • упрек
	Эмоциональные жанры: <ul style="list-style-type: none"> • сочувствие • утешение
	Конфликт: <ul style="list-style-type: none"> • угроза • ссоры • споры

	<ul style="list-style-type: none"> • оскорбления • дразнилки
	Юмор: <ul style="list-style-type: none"> • шутка • розыгрыш • анекдот • насмешка • колкость
	Рассказ: <ul style="list-style-type: none"> • ответы на реплики взрослых • рассказ об отсутствующем лице • монологи-рассказы • рассказ-фантазия
	Письменные Жанры интернет общения: <ul style="list-style-type: none"> • SMS • переписка в Интернет-среде • электронный дневник • комментарии к фото • статус в блоге Письменные тексты: <ul style="list-style-type: none"> • бытовая записка • дневник • анкета • поздравительная открытка
	Школьный дискурс: <ul style="list-style-type: none"> • пересказ • доклад • презентация • сочинение/изложение • ответ на уроке

Allikas: Tabel on koostatud TÜ Narva Kolledži vene keele lektoraadi esindajate ja tudengite poolt.

Tabel 3. Otsingu kriteeriumid.

Параметры поиска	
возраст ребенка	одно значение
пол ребенка	одно значение
подкорпус (устный/ письменный)	одно значение
тип речи (нарратив/ рассуждение/ описание)	множественный выбор
разновидность речи (репродуктивный/ продуктивный/ продуктивно-репродуктивный)	одно значение

Allikas: Tabel on koostatud lõputöö autori poolt TÜ Narva Kolledži vene keele lektoraadi informatsiooni alusel

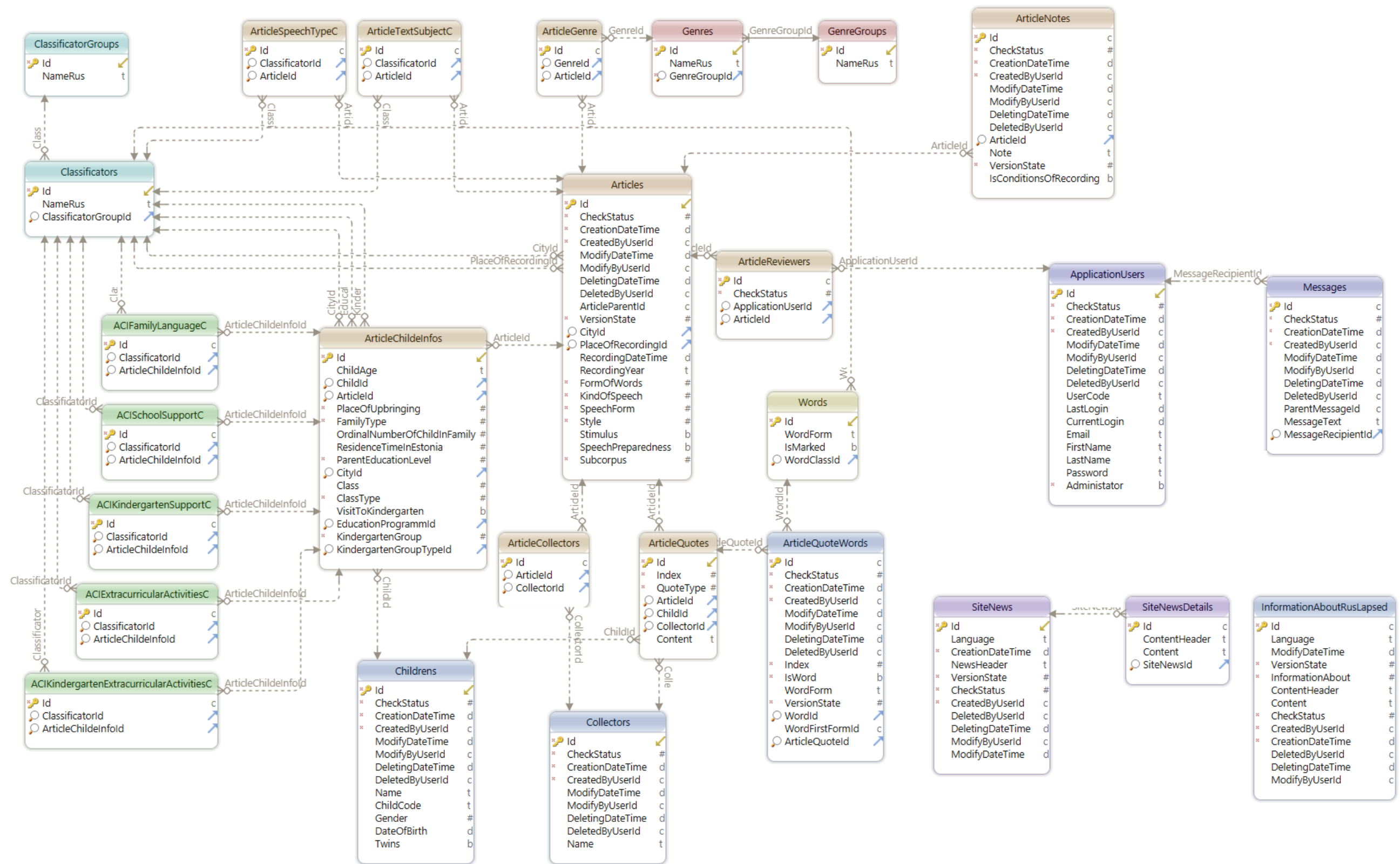
Tabel 4. Teksti vormistuse standart.

Условный знак	Значение	Фрагмент текста
< >	1) информация о ситуации записи текста; 2) метатекстовые включения (смех, кашель и т. п.);	1) <Ребенок получил задание пересказать сюжет мультфильма.> 2) ...Мишка варенье собирал / а потом у него лапа в банке застряла <смеется>...
[Алиса, 5,7:]	1) вымышленное имя ребенка; 2) возраст (год, месяц)	
[Собиратель:]	далее следует речь собирателя	
/	границы синтагмы	держит в руках сачок / ловит {mmm} гусеничку маленькую /
. ? !	границы фразы	Нет. Нет / полечу! Кто / такой / вот / рычит и злится / кто?
«...»	прямая речь до и после авторских слов	«Ну / и какие это звери?» — сказала мама.
—	прямая речь, начинающаяся с нового абзаца.	Мышка вышла гулять / и пошла к маме / и сказала маме /:

		— Мама / я видела двух зверей / один страшный другой/ хороший. — Ну/ и какие это звери? — сказала мама.
[неразб]	неразборчивое произнесение слова	Пичешечка [неразб]
слово [=«перевод»]	«перевод» слова в искажен- ной фонетической форме на литературный русский язык	навица [=нравится]
{ }	пауза хезитации	Маша / держит в руках сачок / ловит {mmm} гусеничку маленькую /
[...]	долгая пауза в речи	И еле мальчик сам вылез. [...] {М-м} / и пото [неразб] / [...]

Allikas: Tabel on koostatud TÜ Narva Kolledži vene keele lektoraadi esindajate ja tudengite poolt.

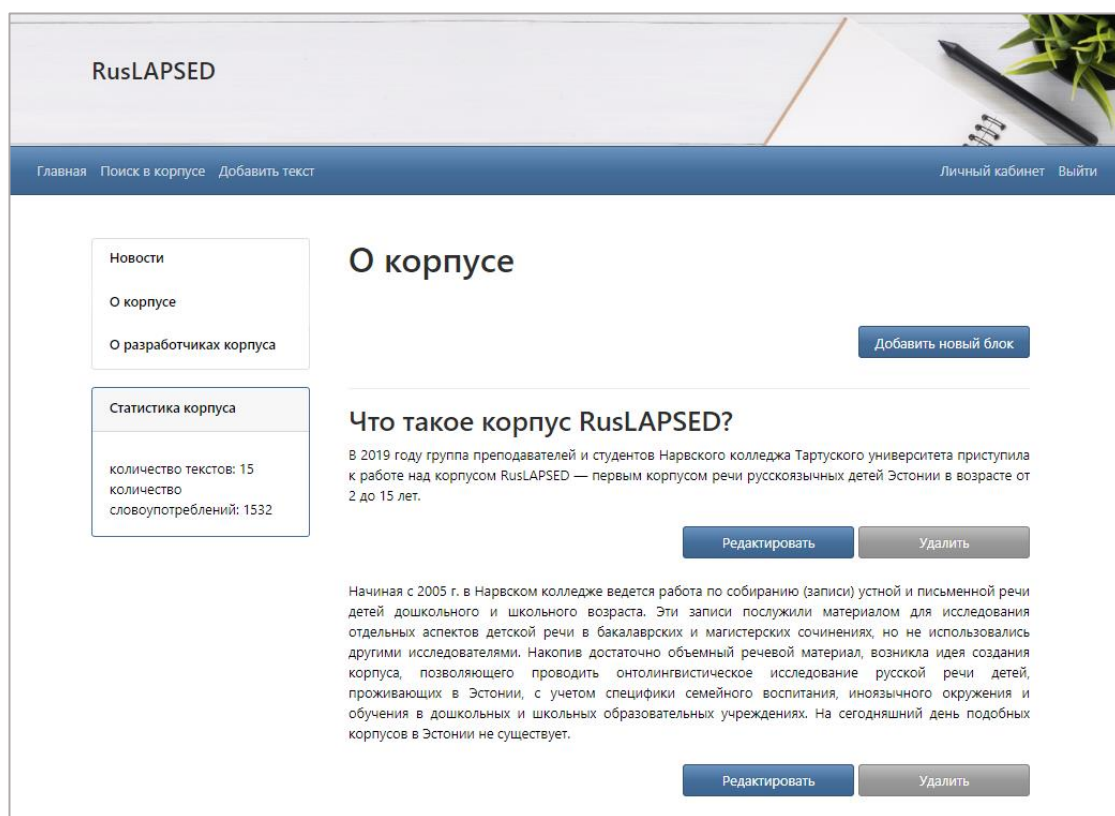
LISA 6. RusLAPSED andmemuudel



Andmed: Autori joonis.

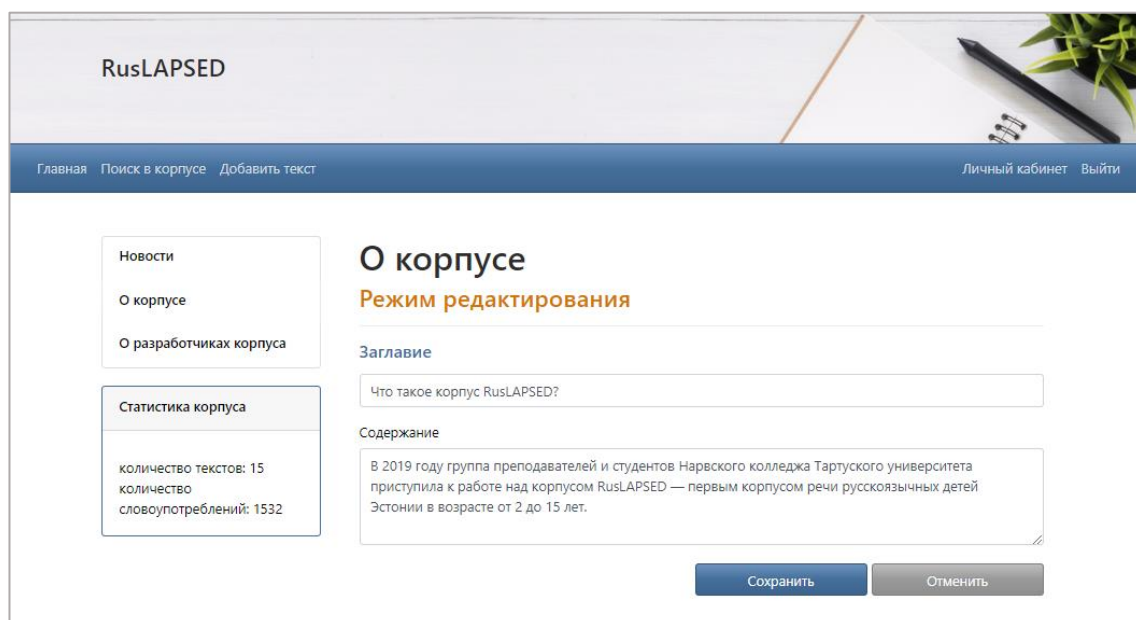
Joonis 1. Andmemudel.

LISA 5. RusLAPSED kasutajaliidese pildid



Andmed: Autori ekraanipilt.

Joonis 1. Avaleht, korpuse informatsiooni moodul, vaade administraatori rollis.



Andmed: Autori ekraanipilt.

Joonis 2. Avaleht, korpuse informatsiooni muutmise/lisamise moodul, vaade administraatori rollis.

RusLAPSED

Главная Поиск в корпусе Добавить текст Личный кабинет Выйти

Новости

О корпусе

О разработчиках корпуса

Статистика корпуса

количество текстов: 15
количество
словоупотреблений: 1532

Новости

Режим редактирования

Тема новости или основное заглавие

Новостной блок

Заглавие

Содержание

Добавить новый блок

Сохранить Отменить

Andmed: Autori ekraanipilt.

Joonis 3. *Avaleht, korpuse uudise muutmise/lisamise moodul, vaade administraatori rollis.*

RusLAPSED

Главная Поиск в корпусе Добавить текст Личный кабинет Выйти

Поиск Фильтры Очистить фильтры Поиск

Найдено текстов: 15

Тексты

Показать только мои тексты

Дата добавления текста: 18.05.2020
Количество детей, принимавших участие в записи: 1

<Ребенок сочинил стихотворение.>

[Ярослав , 10,2:]

Зима

[Ярослав , 10,2:]

Придёт зима, ударит морозом. ...

Детальный просмотр

Andmed: Autori ekraanipilt.

Joonis 4. *Teksti otsingu leht, vaade administraatori rollis.*

RusLAPSED

[Главная](#)
[Поиск в корпусе](#)
[Добавить текст](#)

[Личный кабинет](#)
[Выйти](#)

Дата добавления текста: 18.05.2020

Количество детей, принимавших участие в записи: 1

Количество словоупотреблений: 20

<Ребенок сочинил стихотворение.>

[Ярослав , 10,2:]

Зима

[Ярослав , 10,2:]

Придёт зима, ударит морозом.

[Ярослав , 10,2:]

Придёт зима, золото скинет.

[Ярослав , 10,2:]

Придёт зима, деда [=Деда] Мороза разбудит

[Ярослав , 10,2:]

А уйдёт зима, и весна поступит [=наступит].

▼ ИНФОРМАЦИЯ О ТЕКСТЕ

▼ Стандарт оформления текста

Редактировать

Назад

Andmed: Autori ekraanipilt.

Joonis 5. Täisteksti kuvamine, vaade administraatori rollis.

^ ИНФОРМАЦИЯ О ТЕКСТЕ

Общая информация

Собиратель	Людмила Николаева	Форма речи	письменная
Дети	(Код: 22) Ярослав	Вид речи	продуктивная
Год записи	2019	Стиль	нейтральный
		Жанр	стихотворение
		Тематика текста	детский сад времена года
		Подкорпус	письменный

▼ Стандарт оформления текста

Andmed: Autori ekraanipilt.

Joonis 6. Täisteksti kuvamise mooduli osa, teksti metaandmed, vaade administraatori rollis.

^
ИНФОРМАЦИЯ О ТЕКСТЕ

Информация о ребенке

Имя	Ярослав	Класс	3
Код	22	Тип класса	языковое погружение
Пол	♂	Посещение ДДУ	нет
Дата рождения	03.07.2009		
Близнец(ы)	нет		
Возраст ребенка на момент записи	10,2		
Место воспитания	семья		
Тип семьи	полная		
Номер ребенка в семье	1		
Продолжительность жизни в Эстонии	37		
Уровень образования родителей	высшее		

Назад

v
Стандарт оформления текста

Andmed: Autori ekraanipilt.

Joonis 7. Täisteksti kuvamise mooduli osa, lapse metaandmed, vaade administraatori rollis.

v
ИНФОРМАЦИЯ О ТЕКСТЕ

^
Стандарт оформления текста

Условный знак	Значение	Фрагмент текста
< >	метатекстовые включения (смех, кашель и т. п.)	...Мишка варенье собирал / а потом у него лапа в банке застряла < смеется > ...
[Алиса, 5,7:]	1) вымышленное имя ребенка 2) возраст (год, месяц)	
[Собиратель:]	далее следует речь собирателя	
/	границы синтагмы	держит в руках сачок / ловит (mmm) гусеничку маленькую /
. ? !	границы фразы	Нет. Нет / полечу! Кто / такой / вот / рычит и злится / кто?
«...»	прямая речь до и после авторских слов	«Ну / и какие это звери?» — сказала мама.

Andmed: Autori ekraanipilt.

Joonis 8. Täisteksti kuvamise mooduli osa, teksti vormistuse reeglite osa, vaade administraatori rollis.

Дата добавления текста: 18.05.2020
Количество детей, принимавших участие в записи: 1
Количество словоупотреблений: 20

<Ребенок сочинил стихотворение.>

[Ярослав , 10,2:]
Зима

Словоформа

Зима

Часть речи

Редактировать

Разметка метаданных словоформы

Удалить

Andmed: Autori ekraanipilt.

Joonis 9. Täisteksti kuvamise mooduli osa, sõna lisaandmetega vaatamise režiim, vaade administraatori rollis.

RusLAPSED

Главная Поиск в корпусе Добавить текст Личный кабинет Выйти

Разметка метаданных словоформы

Словоформа

ЗИМА

Часть речи

Сохранить

Назад

Andmed: Autori ekraanipilt.

Joonis 10. Sõna märgistuse leht, vaade administraatori rollis.

RusLAPSED

Главная Поиск в корпусе Добавить текст Личный кабинет Выйти

Добавить новый текст

Метаразметка текста

▼ СОБИРАТЕЛИ

▼ ДЕТИ

▼ ИНФОРМАЦИЯ О ТЕКСТЕ

Текст

Комментарии, информация о ситуации записи текста

Чтобы добавить текст, укажите собирателя и имя ребенка

▼ ИНФОРМАЦИЯ О ТЕКСТЕ

▼ Стандарт оформления текста

Сохранить текст

Andmed: Autori ekraanipilt.

Joonis 11. Teksti lisamise leht, vaade administraatori rollis.

▲ СОБИРАТЕЛИ

Выбрать собирателя из базы данных

Поиск по имени

Имя собирателя

Имя собирателя

Людмила Николаева

Добавить нового собирателя

Добавить собирателя

Andmed: Autori ekraanipilt.

Joonis 12. Teksti lisamise leht, koguja lisamine, vaade administraatori rollis.

СОБИРАТЕЛИ

Людмила Николаева

ДЕТИ

Выбрать ребенка из базы данных

Поиск по имени

Поиск по коду

Код и имя ребенка

22 - Ярослав

ИНФОРМАЦИЯ О РЕБЕНКЕ

Имя	Ярослав
Код	22
Пол	
Дата рождения	03.07.2009
Близнец(ы)	нет

Andmed: Autori ekraanipilt.

Joonis 13. Teksti lisamise leht, lapse lisamine, vaade administraatori rollis.

ИНФОРМАЦИЯ О СЕМЬЕ

Возраст ребенка

Язык(и) семейного общения

русский — эстонский —

русский

+

Добавить язык

Место воспитания

☐ семья
☐ дом ребенка

Тип семьи

☐ полная
☐ неполная

Номер ребенка в семье

Место жительства семьи

Йыхви

+

Добавить город

Продолжительность жизни в Эстонии

Уровень образования родителей

☐ основное
☐ среднее
☐ среднее специальное
☐ высшее

УСЛОВИЯ ОБУЧЕНИЯ

☐ детское дошкольное учреждение
☒ школа

ШКОЛА

Класс

Тип класса

☐ обычный
☐ языковое погружение
☐ специальный

Посещение ДДУ

☐ да
☐ нет

Программа обучения

государственная

+

Добавить программу обучения

Поддержка, оказываемая в учебном заведении

логопед

+

Добавить вид поддержки, оказываемой в учебном заведении

Внешкольная учебная деятельность

занятия с репетитором по другим предметам

+

Добавить внешкольную учебную деятельность

Добавить нового ребенка

▼

Добавить ребенка

Andmed: Autori ekraanipilt.

Joonis 14. Teksti lisamise leht, lapse metaandmete märgistamine, vaade administraatori rollis.

▼ СОБИРАТЕЛИ

Людмила Николаева

▼ ДЕТИ

Ярослав

▲ ИНФОРМАЦИЯ О ТЕКСТЕ

Условия записи

Место записи

Йыхви

Добавить город

школа

Добавить место записи

Дата записи

mm/dd/yyyy

Год записи

Текст

Форма речи

☒ устная

☐ письменная

Разновидность речи

☐ монолог

☒ диалог

☐ полилог

Вид речи

☒ продуктивная

☐ репродуктивная

☐ продуктивно-репродуктивная

Тип речи

описание

Добавить тип речи

Стиль

☐ нейтральный

☐ сниженный

Жанр

разговор

ролевой диалог в сюжетной игре

Добавить жанр

Тематика текста

времена года

Добавить тематику текста

Стимул

☒ да

☐ нет

Степень подготовленности речи

☐ подготовленная

☐ неподготовленная

Подкорпус

☒ устный

☐ письменный

Добавить информацию о тексте

Andmed: Autori ekraanipilt.

Joonis 15. Teksti lisamise leht, teksti metaandmete märgistamine, vaade administraatori rollis.

RusLAPSED

Главная
Поиск в корпусе
Добавить текст

Личный кабинет
Выйти

Личные данные

Персональный код в инфосистеме	
Адрес электронной почты	olga.gandshu@xxx.xxx
Имя	Olga
Фамилия	Gandshu
Дата и время последнего входа	19.05.2020, 13:12
Роль пользователя в системе	Администратор
Статус	Действующий

Выслать новый пароль
Заблокировать
Назад

Andmed: Autori ekraanipilt.

Joonis 16. Kasutaja leht, isikliku andmete moodul, vaade administraatori rollis.

RusLAPSED

Главная
Поиск в корпусе
Добавить текст

Личный кабинет
Выйти

Личные данные

Персональный код в инфосистеме

Дата и время последнего входа
19.05.2020, 11:58

Роль пользователя в системе
Администратор

Имя пользователя и адрес электронной почты

olga.gandshu@

Данный адрес электронной почты используется для входа в систему и восстановления пароля.

Имя
Olga

Фамилия
Gandshu

Сохранить
Назад

Andmed: Autori ekraanipilt.

Joonis 17. Kasutaja leht, isikliku andmete muutmise moodul, vaade administraatori rollis.

RusLAPSED

Главная Поиск в корпусе Добавить текст Личный кабинет Выйти

Смена пароля

Новый пароль должен содержать минимум восемь символов, как минимум одну заглавную букву, одну строчную букву и одну цифру

Текущий пароль

Новый пароль ❗
Пароль не отвечает требованиям

Подтверждение нового пароля ❗
Пароли не совпадают

Сохранить Назад

Andmed: Autori ekraanipilt.

Joonis 18. Kasutaja leht, parooli muutmise moodul, vaade administraatori rollis.

RusLAPSED

Главная Поиск в корпусе Добавить текст Личный кабинет Выйти

Добавление нового пользователя

На указанный вами адрес электронной почты будет отправлено письмо с информацией о регистрации в инфосистеме RusLAPSED

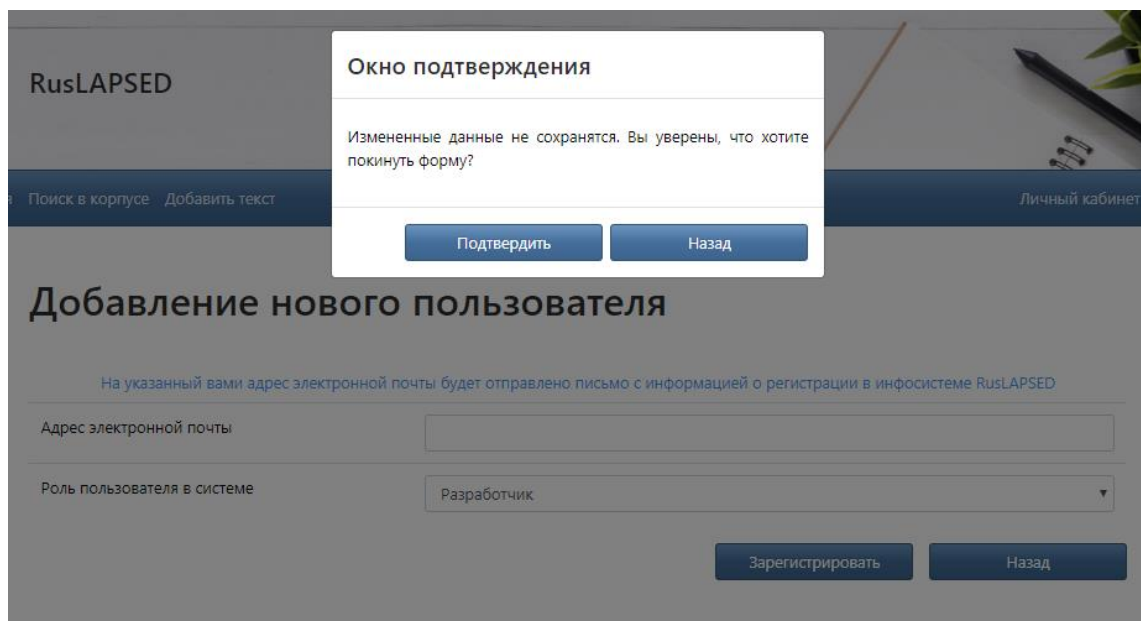
Адрес электронной почты

Роль пользователя в системе

Зарегистрировать Назад

Andmed: Autori ekraanipilt.

Joonis 19. Kasutaja leht, uue kasutaja registreerimise moodul, vaade administraatori rollis.



Andmed: Autori ekraanipilt.

Joonis 20. Kinnituse aken, vaade administraatori rollis.

LOGO RusLAPSED

Добро пожаловать, Ольга!

Выход

Главная	О проекте	Добавить текст	Моя страница	Настройки
---------	-----------	----------------	--------------	-----------

Поиск текста по слову

Использовать фильтр

Найти

Строка с информацией о количестве найденных текстов

Фильтрация по метаданным текста

Ребенок

Пол

Возраст

Семья

Язык семейного общения

Место воспитания

Тип семьи

Номер ребенка в семье

Близнецы

Место жительства семьи

Продолжительность жизни в Эстонии

Уровень образования родителей

Условия обучения

Детское дошкольное учреждение ☒
 Группа
 Тип группы

Школа ☐
 Поддержка, оказываемая в учебном заведении
 Внешкольная учебная деятельность

Условия обучения

Детское дошкольное учреждение ☐
 Посещение ДДУ
 Поддержка, оказываемая в учебном заведении

Класс
 Внешкольная учебная деятельность

Тип класса

Программа обучения

Место записи

Детское дошкольное учреждение ☐
 Город

Школа ☒

Текст

Форма речи

Разновидность речи

Вид речи

Тип речи

Стиль

Жанр

Степень мотивированности речи

Степень подготовленности речи

Тематика текста
 Стимул

Фильтрация по метаданным слова

Выбрать часть речи

Подробная фильтрация по метаданным текста показана в отдельном файле

10.11.2019

Lorem ipsum dolor sit amet et delectus accommodare his consul copiosae legendos at vix ad putent delectus delicata usu. Vident dissentiet eos cu eum an brute copiosae hendrerit. Eos erant dolorum an. Per facer affert ut. Mei iisque mentitum moderatus cu...

07.11.2019

Lorem ipsum dolor sit amet et delectus accommodare his consul copiosae legendos at vix ad putent delectus delicata usu. Vident dissentiet eos cu eum an brute copiosae hendrerit. Eos erant dolorum an. Per facer affert ut. Mei iisque mentitum moderatus cu...

<
1
2
3
4
5
6
7
.....
114
115
116
117
118
119
120
>

Joonis 1. Otsingulehe filtritega eskiis.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Olga Gandshu,

1. Annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „LASTE KÕNE KORPUSE „RusLAPSED“ MÄRGISTUSE JA OTSINGU INFOSÜSTEEM“, mille juhendaja on Andre Säask, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Olga Gandshu

21.05.2020